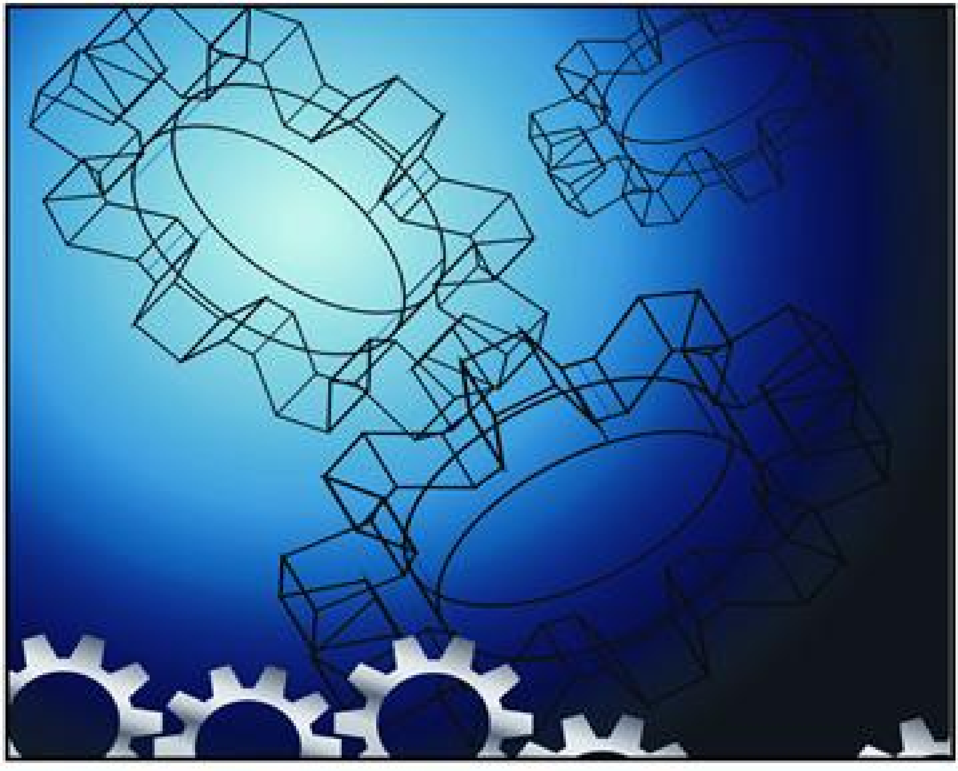
PREMIER REFERENCE SOURCE

MANUFACTURING INTELLIGENCE FOR INDUSTRIAL ENGINEERING

Methods for System Self-Organization, Learning, and Adaptation



Zude Zhou, Huaiqing Wang & Ping Lou

Manufacturing Intelligence for Industrial Engineering:
Methods for System Self-Organization, Learning, and Adaptation

Zude Zhou
Wuhan University of Technology, China

Huaiqing Wang
City University of Hong Kong, Hong Kong

Ping Lou
Wuhan University of Technology, China

ENGINEERING SCIENCE REFERENCE

ENGINEERING SCIENCE REFERENCE

Hershey · New York

Director of Editorial Content: Kristin Klinger
Director of Book Publications: Julia Mosemann
Acquisitions Editor: Lindsay Johnston
Development Editor: Joel Gamon
Publishing Assistant: Deanna Zombro
Typesetter: Michael Brehm
Production Editor: Jamie Snavely
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Engineering Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

Library of Congress Cataloging-in-Publication Data

Zhou, Zude, 1946-
Manufacturing intelligence for industrial engineering : methods for system self-organization, learning, and adaptation / by Zude Zhou, Huaiqing Wang, and Ping Lou.
p. cm.
Includes bibliographical references and index.
Summary: "This book focuses on the latest innovations in the process of manufacturing in engineering"--Provided by publisher.
ISBN 978-1-60566-864-2 (hardcover) -- ISBN 978-1-60566-865-9 (ebook) 1. Technological innovations. 2. Industrial engineering. 3. Artificial intelligence. I. Wang, Huaiqing. II. Lou, Ping. III. Title. T173.8.Z486 2010
670.285--dc22
2009034472

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Table of Contents

Foreword

Manufacturing engineering has come a long way, from the "black art" in the 1800s to the first scientific analysis of machining operations by F.W. Taylor in early 1900s (On the Art of Cutting Metals, 1906). In the early 1950s, computers were developed to take control of machine tools and NC machines were born, and later, CNC machines. The 60s and 70s saw a rapid proliferation of software and hardware development in support of manufacturing operations in the form of design, analysis, planning, processing, measurement, dispatch and distribution. The late M Eugene Merchant, then Director of Research Planning of Cincinnati Milacron Inc., made an exciting Delphi-type technological forecast of the future of production engineering at the General Assembly of CIRP in Warsaw, 1971. Five years later, he made another report on the "Future Trends in Manufacturing – Towards the Year 2000" in the 1976 CIRP GA in Paris. He reported that between then (1976) and the year 2000, the overall future trend in manufacturing will be towards the implementation of the computer-integrated automatic factories. More than 30 years had since whisked past, manufacturing technologies had indeed progressed even more rapidly than Dr Merchant's prediction then.

Manufacturing operations have changed from programmed operations to programmable operations. In the last two decades, many manufacturing operations and processes have become near autonomous, i.e. they possess sufficient intelligence to diagnose, optimize, decide and correct any actions with minimum human interaction. Some systems can acquire and learn from past cases and become increasingly more "learned" through usage. Machine tools which are Internet-enabled can be continuously monitored by their manufacturers and their "state-of-heath" is exactly known and predictable to enable the reduction of breakdown time and to ensure timely maintenance. Computer-integrated Manufacturing (CIM) has evolved to become Computer-Human Integrated Manufacturing (CHIM). Seamless integration of human and computer intelligence is another measure to capture the perfect complementation between man and machine.

It is with great pleasure to witness this new book '*Manufacturing Intelligence for Industrial Engineering: Methods for System Self-Organization, Learning and Adaption'* by Zude Zhou, Qinghuai Wang and Ping Lou. It is a timely capture of the state-of-the-art development of intelligent manufacturing processes, covering a vast amount of materials from design, planning, diagnosis, information control, agents, and many enabling platforms and supporting theories. I have, beyond doubt, that this contribution will be invaluable to researchers as well graduate students in the field of manufacturing engineering.

I sincerely congratulate the authors on having produced this splendid new book

A. Y. C. Nee, DEng, PhD
National University of Singapore
Regional Editor IJAMT
Regional Editor IJMTM

A. Y. C. Nee *received his PhD from the Victoria University of Manchester in 1973 and Doctor of Engineering (DEng) degree from UMIST in 2002. He joined then University of Singapore as a faculty member in 1974. He has held various administrative positions including Head of Department of Mechanical Engineering from 1993 to 1996, Dean of Faculty of Engineering from 1995 to 1998, other appointments include: Director of Office of Quality Management, Dean of Admissions, CEO of Design Technology Institute, Co-Director Singapore-MIT Alliance, Deputy Executive Director, then NSTB SERC, Director of Office of Research. Prof Nee received his National Day Award in Public Administration—PPA(P) in 2007. Professor Nee is well known in the field of manufacturing engineering. His research focuses on computer-aided design of fixtures, molds and dies, distributed manufacturing systems, AI and augmented reality applications in manufacturing. He was selected a Fellow of the Society of Manufacturing Engineers with citation in 1990, and a Fellow of the International Academy for Production Engineering (CIRP) in the same year. He was elected as Vice-President (Elect) at the CIRP recent senate meeting in August 2009, and will be Vice President in August 2010 and President of CIRP from August 2011. He has published over 250 papers in international refereed journals, 5 authored and 5 edited books. Professor Nee is regional editor of International Journal of Machine Tools and Manufacture, and International Journal of Advanced Manufacturing Technology. In addition, he is editorial board member and associate editor of another 20 refereed journals. He is also Chairman of an NUS spin-off company—Manusoft Technologies Pte Ltd established in 1997.*

Preface

The environment of the manufacturing industry has changed impressively during this half century. New theories and technologies in the field of computers, networks, distributed computation, and artificial intelligence are extensively used in the manufacturing area. Integration and intelligence have become the developing trends of future manufacturing systems. These inform the concept of manufacturing change from the narrow sense of fabrication technique to the broad sense of extensive manufacture, that is, from the transformation of raw materials into finished goods, to the whole process of the product life cycle involving product design, fabrication, planning, managing, and distribution. Intelligent manufacturing will become one the most promising manufacturing technologies in the next generation of manufacturing industries.

Manufacturing Intelligence (MI), as a new discipline of manufacturing engineering, focuses on scientific foundations and key technologies for developing, describing, integrating, sharing, and processing intelligent activities in the process of manufacturing. It mainly covers intelligent-control theory and technology for manufacturing equipment, intelligent management and decision making for the manufacturing process, intelligent processing of manufacturing information, representation and reasoning of manufacturing knowledge, as well as intelligent surveillance and diagnosis for manufacturing equipment and systems.

Clearly, MI is different from Artificial Intelligence (AI). AI is one aspect of theoretical research led by the requirements of mimicking human intelligence. It mainly focuses on exploring the mechanism of the process of human intelligent activities and emphasizes general theories, which highlight explorations of theory, as well as having serious logicality and reasoning. By contrast, MI mainly studies the mimicry of human intelligence to solve issues with intelligent computers (including software and hardware), and is a type of foundational research led by the requirements of applications in the manufacturing field. Although these two disciplines are different, they are related each other. AI is one of the main foundations of MI and the development of MI and the solution to the issues unsolved by AI will accelerate the development of AI.

This book consists of four parts with fourteen chapters which include engineering background, foundations, technologies, applications, implementations, case studies, trends of intelligent manufacturing, and prospects for manufacturing intelligence. Part I contains one chapters, *viz*. chapter 1, which introduces manufacturing intelligence, the development of intelligent manufacturing, and the features of intelligent activities in the process of manufacturing. Part II and Part III including twelve chapters constitute the main part of this book. In these two parts, scientific foundations, key technologies and pragmatic applications of manufacturing intelligence are analyzed. Among them, chapters 2 to 8 composing the Part II offer an extensive presentation of the engineering scientific foundations in manufacturing intelligence. Chapter 2 describes knowledge-based systems which mainly details general approaches for knowledge representation, acquirement, and general techniques for searching and reasoning. Chapter 3 presents

an overview of intelligent agents and multi-agent systems. Chapter 4 contains the principle and techniques of data mining and knowledge discovering. Chapter 5 introduces the principle and applications of computational intelligence in engineering and manufacturing, including neural networks, genetic algorithms, and fuzzy logic. Chapter 6 has an overview of information system modeling, including the general processes and strategies, some different modeling approaches and modeling tools. Chapter 7 includes an overview of multi-sensor integration and data fusion theories. Chapter 8 introduces the principle and approaches to group theory, including coding systems for parts, approaches for grouping parts and applications in manufacturing designing and processing. Chapters 9 to 13 make up Part III of the book: the applications and case studies for manufacturing intelligence. Chapter 9 presents the structure theory of intelligent control, a general architecture of the intelligent controller, and intelligent systems. Chapter 10 contains knowledge-based approaches for designing, beginning with the basic concepts and approaches of conventional computer-aided design (CAD) systems. Chapter 11 includes an overview of computer-aided process planning, including concepts and enabling technologies, and the architecture and decision-making process of intelligent computer-aided process planning is also presented. Chapter 12 presents an overview of remote monitoring and intelligent diagnosis. Chapter 13 consists of the principles and approaches to intelligent management and decision-making in manufacturing. Like Part I, Part VI also contains only one chapter, *viz*. chapter 14. In chapter 14, first the summarization of the theories, technologies, and applications in the aforementioned chapters is presented, and then these intelligent manufacturing technologies compare to the traditional manufacturing technologies. Last the prospects for manufacturing intelligence and the trends of intelligent manufacturing in the future are discussed..

This book is intended primarily for senior undergraduate and graduate students in mechanical, electromechanical and industrial engineering programs. Its integrated treatment of the subject makes it a suitable reference for practicing engineers and other professionals who are interested in pursuing research and development in this field. For professors and students, this book may be used for teaching as well as self-study. It gives them an up-to-date, in-depth source of material on manufacturing intelligence. For researchers, the publication helps them better understand the field as a whole. They will obtain valuable enlightenment for their future research activities.

The book also provides readers with the scientific foundations, theories, and key technologies of manufacturing intelligence. Hence, readers may use this publication achieve two different but overlapping goals. Firstly, it may help readers to understand manufacturing intelligence in a deeper and more comprehensive way. Furthermore, throughout this book numerous references to literature sources are provided, enabling interested readers to further pursue specific aspects of manufacturing intelligence.

Xue Ligong, Jiang Xuemei, Zhang Xiaomei, Liu Hong, Wang Sheng, Ai Qingsong and Ming Hui compiled the various chapters. I wish to extend my thanks to them for their fruitful work.

The book is supported by the International Cooperation Key Project (Multi-agent based digital manufacturing new theory and new method, grant no. 2006DFA73180) from the Scientific and Technology Committee of China.

Chapter 1
Intelligent Manufacturing and Manufacturing Intelligence

Manufacturing is a prime generator of wealth and is critical in establishing a sound basis for economic growth. Manufacturing is also the cornerstone of all economic activities, and efforts to continuously advance manufacturing technology are therefore vital to a richer and more stable future. Intelligent Manufacturing (IM), believed to be the next generation advanced manufacturing paradigm is extensively investigated by industry and academia. In this chapter, we firstly recall the course of manufacturing development and summarize the characteristics of the four revolutions in this course. Subsequently, the broad sense of 'manufacturing' is articulated and the characteristics of manufacturing activities in the operation of manufacturing processes are depicted. The differences and relationships between Artificial Intelligence (AI) and Manufacturing Intelligence (MI) are then presented. The background of intelligent manufacturing, the attributes of intelligent manufacturing technology and the future development of intelligent manufacturing system are described. Lastly, a summary of this chapter is given.

DOI: 10.4018/978-1-60566-864-2.ch001

INTRODUCTION

Manufacturing has played, and continues to play, a vital role in the world economy. In recent years, manufacturing has undergone profound changes because of the development of science and technology, the requirements of global manufacturing and the changing manufacturing environment. Changes have to be made in order to satisfy the increasingly changing and diversified demands of customers. These changes are bringing manufacturing from a resource-based centralized paradigm to a knowledge-intensive, innovation-based, adaptive, digital and networked one. Integration and intelligence are two vital factors of modern manufacturing. Looking at the history and the present state of manufacturing, it is clear that there have been four revolutions according to the four stages of manufacturing industrial development (Wang, 2005). These are the age of craftsmanship, the age of machines and hard automation, the age of information and flexible automation, and the age of knowledge and intelligent automation.

In the age of craftsmanship, all manufacturing activities from raw materials to finished products

were entirely performed by physical labor, in which a person with hand tools was used to make objects. The quality of products relied very much on individual skills. As technology progressed, such as animal, wind and water power were gradually employed, and more sophisticated tools were developed, but the basic structure of craft-based production remained unchanged.

The industrial revolution that dates back 200 years brought manufacturing to the age of machines and hard automation, and machinery has played an increasingly prominent role since then. In the early days, the mechanization of manual procedures was the first step towards automation. In the later stage of the period of mechanization, the total process of production was analyzed and subdivided into a number of simpler production functions as products were becoming increasingly complex. Workers were carefully but rather narrowly trained to operate their own tools and specialized machines. Such a manufacturing pattern is well suited to mass production.

In the age of information and flexible automation, information processed by computers and automatic control has led to significant changes in manufacturing patterns and technologies. At the early stage of this period, a great deal of emphasis was placed on the development and application of 'hardware' and on the search for hard automation. In the later stage of this period, information and flexible automation have been the primary focus of development. In order to improve production efficiency, considerable effort has been placed on the development and application of new manufacturing techniques in programmable equipment (hardware: such as NC, CNC, and robotics,) and the computer-aided systems (software: such as CAD, CAPP, CAE, CAM, and so forth).

In today's information age, manufacturing is forced to be more intelligent in order to have the power to process escalating manufacturing information and to satisfy dynamical marketing demands. With the rapid development of Artificial Intelligence (AI) and the quick improvement various related enabling technologies, the manufacturing age of knowledge and intelligent automation has already begun. The main focus in this age will be on manufacturing flexibility and adaptability in the various aspects of manufacturing, such as automatic and intelligent design, production planning and control, configuration management, intelligent decision support, automatic and intelligent failure detection and maintenance, and so on. Furthermore, Manufacturing systems' self-organization, self-learning and adaptation according to the conditions of outside environments will also be of paramount importance. One of the developing trends of future manufacturing systems will be intelligent and knowledge-intensive systems, which focus on the integration of knowledge available from various manufacturing domains and the combination of human-computer intelligence.

MANUFACTURING ACTIVITIES

The word 'manufacturing' includes much more than the basic fabrication techniques. It involves diverse activities from raw materials to finished products during the processes of manufacturing, including marketing prediction, procurement, design, plan, fabrication, distribution, as well as recycling and services. Customer needs or marketing predictions are the beginning of manufacturing activities. Products are designed according to customer needs. Product design is a complicated process, involving conceptual design, configuration, parametrization, and manufacturable analysis. Product design is normally made accessible to a planning subsystem, including process planning, scheduling and manufacturing resource planning, which transforms product design into production plan and manufacturing resource plan. After the plans are developed, the product can then be manufactured. With the rapid development of manufacturing intelligence and information technology, the process of pro-

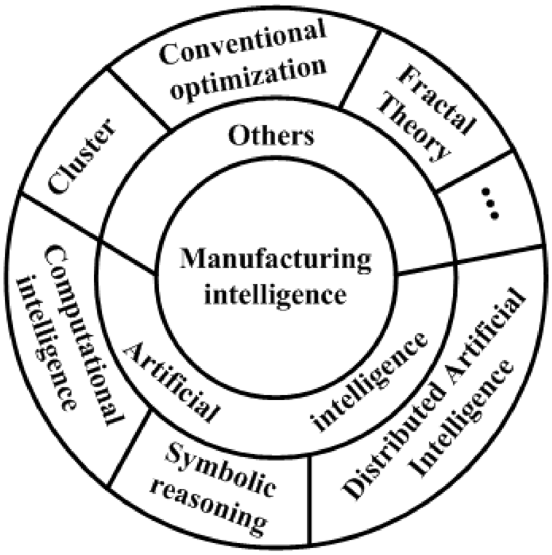
duction could usually be made autonomous or semi-autonomous. To ensure that the process is under control, it is necessary to monitor the process and obtain status information about the different processes and product variables. If there happen to be faults, the functions of diagnosis and maintenance start working to make instigate recovery. The principal functionalities of production are process planning, scheduling, material resource planning, monitoring, diagnosis, control, inspection, and assembly. Each subsystem, such as design, planning, production, and so forth functions effectively as the fundamental objective of optimizing the running of manufacturing systems as a whole. For manufacturing systems to run optimally, however, they are not only dependent on the integration of each subsystem, but also the cooperation and coordination of each subsystem. In order to ensure that different subsystems cooperate and coordinate, there is a system-level subsystem whose functions are to develop system organization, control strategies, and cooperative mechanism.

ARTIFICIAL INTELLIGENCE AND MANFUACTURING INTELLIGENCE

Artificial intelligence (AI), an expression coined by Professor John McCarthy from Stanford University in 1956, is a branch of computer science concerned with making computers behave like humans by modeling human thoughts on computers. Computational Intelligence (CI), as a new development paradigm of intelligent systems, has resulted from a synergy among Neural Networks (NN), Fuzzy Sets (FS), and Genetic Algorithms (GA) (Engelbrecht, 2007). CI, as one important branch of AI, is an effective complement to AI and also an important component of AI. With the development and application of AI intelligence can be built into machines and manufacturing processes. In general, AI can be divided into two categories: the first is Symbol Reasoning (SR) including Expert Systems (ES), Knowledge-Based Systems (KBS), Case-Based Reasoning (CBR), and the second is CI including NN, FS, and GA. Combining symbolic reasoning with computational intelligence, we can take full advantage of the 'intelligence' provided by computational method and also the logical reasoning power based on knowledge of symbolic reasoning. Hence, a powerful intelligent system can be developed via the combination of SR and CI. When AI, as the core technique of IM, is applied to different manufacturing stages, such as design, planning, operation, quality control, maintenance, marketing and distribution, and so forth, and we regard this as manufacturing intelligence (MI). AI is an indispensable and greatly important fundamental element of MI, and MI is an application of AI in manufacturing. However all techniques, besides AI, which can help improve the intelligence level of manufacturing, are also important components of MI. They include conventional optimization methods, cluster analysis tools, fractal theory, and so on. MI is an extension of AI in the manufacturing domain and also a comprehensive application of diverse non-intelligent optimization techniques in improving automation and intelligence of manufacturing (see Figure 1). Fractal Theory, for instance, provides us with the mathematical tools to analyze the geometrical complexity of natural and artificial objects (Castillo & Melin, 2003); Cluster Analysis provides us with the mathematical tools to classify data according to similarity (Mirkin, 2005); conventional optimization approaches provide us with mathematical tools for optimization problems (Saravanan, 2006).

MI, as a fundament of intelligent manufacturing, is a combination of diverse AI techniques, such as reasoning, learning, self-improvement, goal-seeking, self-maintenance, problem-solving and adaptability, and other available non-intelligent techniques in manufacturing, which exhibit capabilities when applied to solve manufacturing engineering problems in design, planning, production, process modeling, monitoring, inspection,

Figure 1. Manufacturing intelligence



diagnosis, maintenance, assembly, system modeling, control and integration. With the development of related techniques, especially of AI, MI also continuously develops and improves to satisfy the requirements of manufacturing development. In the early stage of development many manufacturing systems improve their intelligence level through the application of knowledge-based ES, and the trait of this stage is symbolic reasoning. Because most ES are closed intelligent systems with weak self-learning abilities, they can only deal with single domain knowledge in a standalone manner, and normally do not communicate with the outside environment. Their capability to handle complex numerical calculation is very limited. However, manufacturing is a complicated process, which needs more than just logical reasoning based on a certain domain, and it also needs capabilities to optimize, self-learn, self-organize, and adapt so as to optimize and improve various functions during the process of manufacturing. AI combined with CI and other techniques is employed in manufacturing to overcome the drawbacks of ES. These new techniques with the attributes of flexibility, generality and precision combined with the knowledge-based symbolic reasoning of ES will continuously improve the intelligence level of manufacturing and provide optimal solutions with the development of MI.

An alternative view of human intelligence has emerged in 1990s, and it is regarded as the capability of a system to interact with its environment without clearly defined goals, to learn from this interaction and, in an incremental fashion, to both articulate and achieve its goals (Reti & Kumara, 1997). Multi-agent technology represents a promising approach to designing an intelligent system as a cluster of intelligent agents, which have the power to deal with distributed problems. A multi-agent system is a group of loosely connected agents that are autonomous and independent. These agents are capable of communicating with each other, processing received messages and making decisions, and learning from experiences collectively. The overall system performance is not globally optimized but develops through the dynamic interaction of agents. The novelty of this approach is in replacing hierarchical architectures with network configurations in which nodes with advanced communication capacities are capable of negotiating how to achieve specified goals without any centralized control. Another similar term is called 'holon'. A holon is an autonomous, cooperative, and sometimes intelligent entity; it can be made up of other holons. A system of holons cooperating to achieve a goal or objective forms a holarchy; the holarchy defines the basic rules for cooperating among holons and therefore limits their autonomy. A holonic Manufacturing System (HMS) is a holarchy that integrates the entire range of manufacturing activities from raw material purchasing to finished product. HMS is a kind of effective system for IMS.

INTELLIGENT MANUFACTURING

The journey from small-sized production to mass-production, and until to the current mass-cutomization, combined with the advent and

development of advanced network technologies and information technologies, has inevitably led to the coming of manufacturing globalization. In order to be adaptive to the varieties of customer demands and increasingly reduced lead times of products, manufacturing systems have to become more intelligent and automotive so that they have the power to process large quantities of information to improve their responsiveness and to resist outside interference. With the developments and applications of AI, especially combined with related domain manufacturing knowledge acquisition, expression, storing and reasoning, strong foundation has been set for implementing the intelligentization of manufacturing technology and dynamically configuring manufacturing systems based on manufacturing intelligence.

The research of intelligent manufacturing began some 20 years ago. The research and development of intelligent manufacturing greatly benefits from the development and application of AI and the relevant enabling technologies. However, until now no clear definition of Intelligent Manufacturing has been given. Wright and Boune (1988) believed that the objective of intelligent manufacturing (IM) is to produce products by modeling and mimicking expert's knowledge and technician's craft for small-batch production without human intervention via the integration of knowledge engineering, manufacturing software systems, robotic vision and control. Manufacturing is considered in its broadest sense to be more than just the basic fabrication techniques: it involves the entire spectrum of manufacturing processes from raw materials to finished products, including design, processing and production planning, machining, marketing, and so forth. With the development and application of new theory and technology, especially the development and application of AI, IM has a new objective, which is to achieve automatization, intelligentization and integration of the entire manufacturing process, including marketing prediction, production making-decision, design, machining, logistics, and marketing and distribution. Through the replacement and extension of part of human intelligence in a dynamic manufacturing environment, self-organization, self-learning, and adaptation of the entire manufacturing process is implemented. Intelligent manufacturing is a new discipline which applies computer science, artificial intelligence, mechanical engineering and system science to industrial manufacturing processes (Rao, Wang, & Cha, 1993).

As the characteristics of products are increasingly improved, the structures of products are becoming more and more complex and diverse; the manufacturing information which needs to be processed is dramatically increased. The capability, efficiency and scale of manufacturing systems being able to process the information are extensively researched by the academic and the industrial. Advanced manufacturing systems without the capability of processing information will not work efficiently; for instance, flexible manufacturing systems do not run without an information driver.

Manufacturing systems are now undergoing great changes that bring them from energy-driven systems to information-driven ones. Manufacturing systems need to be flexible and intelligent so that they can process information from various resources. Intelligent manufacturing is proposed for such a dynamic changing manufacturing environment. Research on intelligent manufacturing aims at the comprehensive intelligentization of the entire manufacturing process through machine intelligence instead of partial human intelligence in a real life manufacturing system, that is to say, human intelligence and machine intelligence are perfectly combined, integrated and shared, which is called manufacturing intelligence. Intelligent manufacturing usually refers to intelligent manufacturing technology (IMT) and intelligent manufacturing system (IMS).

Intelligent Manufacturing Technology

Intelligent Manufacturing Technology (IMT) is the combination of AI and manufacturing technology. It simulates human expert's knowledge in the manufacturing process via computer, with a highly-flexible and highly-integrated manner in various stages of manufacturing process. IMT merges AI with various optimization techniques and applies in the spectrum of manufacturing processes, and it can perform analysis, reasoning, conceiving and decision-making on manufacturing problems - aiming at replacing or extending part of human intelligence in the manufacturing process and collecting, storing, improving, sharing, inheriting and developing the human expert's intelligence. The related intelligent manufacturing technologies are as follows:

Intelligent Control

In modern manufacturing scenarios, the complex nature of manufacturing processes, the demand for high quality products at ever reducing lead times and the effort to achieve higher profitability is forcing researchers as well as practitioners to attain improved process control of computer numerically controlled (CNC) manufacturing systems (Kumar, *et al*., 2007). Intelligent control for equipment is an important cornerstone for manufacturing intelligentization and automation.

Process control in machining has been traditionally classified into the categories of adaptive control with constraints (ACC), adaptive control with optimization (ACO), and geometric adaptive controller (GAC). ACC systems are mainly aimed at adjusting process variable such as force or power in real time relative to appropriate machining conditions to reduce machining costs and increase machine tool efficiency. These systems are capable of maintaining maximum working conditions in an evolving machining process (Masory & Koren, 1983). Ulsoy and Koren (1989) have defined the objective of ACC systems as appropriate feed rate tuning for controlling the cutting force, which is the maximum cutting force relative to threshold tool breakage conditions. ACO systems (Koren, 1983) have also been developed to gain the control over the machining process parameters such as feed rate, spindle speed or depth of cut for maximizing the process response for attaining better quality products. The primary objective of GAC systems is to maximize the quality of the products in terms of finish operations (Wu, 1986). GAC systems take structural deflection and tool wear as machining constraints for optimizing the finishing quality. These are optimization controls under the ideal circumstances, which are shortage of adaptation and robustness. The controlling strategies and controlling parameters can not be regulated dynamically according to the change of environment.

Intelligent control is a class of control techniques which has come to embrace diverse methodologies combining conventional control theory with emergent AI techniques. Various intelligent control approaches, such as Fuzzy Control and Neural Network for Control, are employed to process control. Fuzzy Control is easy to create and has strong robustness. Neural Network for Control takes full advantage of the traits of NN, such as non-linear mapping, self-learning, adaptation, parallel information processing, associative memory, and perfect fault-tolerance. In addition, intelligent approaches are combined with each other to take full advantage of their respective traits. For example, NN is employed in Fuzzy Control to make it with self-learning ability; Fuzzy Control combined with GA is used to make fuzzy membership functions optimization and fuzzy model evolution; GA applied in Neural Network for Control can optimize and regulate the weight of NN. Combining these intelligent approaches is a developing trend of intelligent control.

Virtual Reality (VR) and Augmented Reality (AR)

Virtual Reality is a high-end user-computer interface that involves real-time simulation and interactions through multiple sensorial channels. These sensorial modalities are visual, auditory, tactile, olfactory, and gustatory (Burdea & Coiffet, 2003). VR has many attributes that are appealing to manufacturing. These include natural multimodal interaction (useful in concept design and personnel training), flexibility (beneficial to design revisions and small-batch production), and remote shared access (useful in ergonomic analysis, product approval, training, and marketing, where multi-disciplinary team are involved).

The term Augmented Reality is used to describe systems that blend computer-generated information with the real environment (Azuma, 1997). This combination can be multisensory and might include the enhancement of an image with virtual annotations, the detection and amplification of sounds, or the use of haptic feed back to increase touch sensing. The prominent attributes are to deal with the combination of real-world and computer-generated data. At present, more AR research is concerned with the use of live video or image which is digitally processed and 'augmented' by the addition of computer-generated graphics. Unlike VR, AR enhances the existing environment rather replacing it. AR is also extensively employed in manufacturing, such as assembly, maintenance, and repair of complex machinery. VR and AR are apparently effective technologies in the assembly domain and cannot be strictly separated; however, VR is often used in the early stage of the life-cycle of an assembly, and AR is predominantly used in the control and maintenance phase.

Intelligent Product Design

Product design is the source and basis of innovation, and is also a guarantee for satisfying various customer demands. Product design activity normally requires intelligence support, involving extensive knowledge of designing domains and the experience of experts. In order to extend and strengthen the functions of conventional computer-aided design (CAD) based on computer graphics, AI and Knowledge Engineering (KE) are employed in this area, and known as intelligent CAD (ICAD). With the help of ICAD to generate, select, and evaluate designs, work load can be decreased, quality can be improved, and profit increased.

Design processes need both mathematical and AI techniques to solve numerical and symbolic problems involving large state spaces, but one distinction between design and problem-solving is that the design process rarely has a right and wrong answer. It usually has feasible or infeasible solutions, especially for product conceptual design where the problem is wrongly-defined and information is incomplete; hence, prior experience and knowledge in the particular problem domain is usually very important. This prevents the application of established optimization methods from being used to solve this particular problem. These characteristics of product design call for new AI techniques. Through the application of expert knowledge, rational reasoning, and effective optimal approaches, the activities or part of the activities in product design are completed by computer instead of human brain. Genetic algorithms, for instance, have been successfully used to generate a feasible search space of an arbitrary size and to reach a near global optimization solution, which is a kind of evolutionary techniques and is an iterative optimization tool; Artificial Neural Networks have been developed to deliver the capability for learning from experience in a given domain; Fuzzy Logic Systems have been applied to enable decision-making based on incomplete parameters; Expert Systems have been employed to help analyze the requirements of product and product design with the help of a knowledge base and rational reasoning mechanism. It is obvious

that no single technique from AI will meet the challenge posed by this design problem alone. The various AI techniques need to be deployed jointly to support product design.

Intelligent Production Planning

Process planning, scheduling, and Material Resource Planning (MRP) are considered as three major items in production planning. The function of process planning is the selection of manufacturing equipment and the arrangement of the sequences of operations in manufacturing equipment; scheduling is an optimization process where limited resources are allocated over time among both parallel and sequential activities; and MRP determines what assemblies must be built and what materials must be procured in order to develop products within a certain time.

Computer-Aided Process Planning (CAPP) based on ES has led the way in the field of CAPP in the last thirty years. These types of CAPP systems can not satisfy the requirements of IM because of theirs monolithic nature. They are shortage of flexibility and extendibility, and not suitable for integration with CAD/CAM. AI, especially computational intelligence (CI) provides underlying techniques to make manufacturing planning more intelligent and flexible. Recently ANN is employed to learn from manufacturing data for discovering useful knowledge in order to make better decisions. FLS and ANN can be combined to solve scheduling problems effectively. Optimization techniques like GA are widely used to identify the best schedule or the best resource planning for a manufacturing system. These techniques are suitable for the intelligentization of CAPP, dynamic production planning, and making MRP.

Intelligent Assembly

The automation of production processes is emphasized at the above; however, the manufacturing of a product is incomplete until its constituent parts have been assembled. The assembly design and planning is the last stage of production and has a great effect on the quality of finished products. Artificial intelligence is employed in the analysis and implementation of assembly processes, *viz*., intelligent assembly. It mainly involves two aspects: the automotive and intelligent implementation of assembly processes and the intelligent integration of design and assembly planning processes.

Robots are important tools in the assembly process. Intelligent control technology and different intelligent sensors employed in robots greatly improve the precision of product assembly and promote the assembling capability of robots.

The activity of assembly design and planning is closely related to product design and production. In a concurrent engineering environment, a computer-based intelligent system for design for assembly (DFA) analysis is also an effective way to reduce cost. With products becoming more complex and highly integrated, DFA analysis becomes more and more important. It can help analyze the ease of assembly/subassembly of the products and make comparison of alternative design easy. Nevertheless, this kind of integration of design and assembly planning processes is complicated, including product design, assembly evaluation and redesign, assembly process planning, design of assembly system and assembly simulation. In order to assist designers in the early stages of a product design, various intelligent techniques are combined to construct an intelligent system for DFA, which provide designers and producers the possibility of assessing and reducing the total production cost. Moreover, virtual reality (VR) and augment reality (AR) technology are employed in virtual assembly systems to provide a more natural and intuitive way to help designers and producers evaluate, analyze and plan the assembly of the products.

Intelligent Monitoring and Diagnosis

In today's highly competitive environment, products with high quality, low cost, and short lead-time are vital factors in obtaining and retaining a favorable market position. In many manufacturing processes, process parameters reflect the running state of the production process; in particular, some key process parameters have a very strong relationship with the quality of finished products. The abnormal changes of these process parameters could result in various types of faulty products. The advent of advanced communication and information technology has provided promising mechanisms and tools to improve the standard of manufacturing by using intelligent control techniques. In this environment, a strong need exists in the manufacturing process for the integration of related machines with monitoring and diagnosis function in real-time. In order to ensure that the manufacturing process is under good control and optimal, it is necessary to monitor the process, obtain process and product information, diagnose problems, control the process parameters, and optimize the results.

E-maintenance of manufacturing equipment is an effective means to provide products with high quality and service. By using Ethernet/Internet, Web-enable and wireless communication technology, e-maintenance provides intelligent prediction (diagnosis, self-maintenance, prognostics, and so on.) to prevent unexpected breakdowns. In these functional models of e-maintenance, various intelligent techniques are developed, which consist of artificial intelligence and expert system, ANNs, data mining, time-series, real-time control, and signal processing (Lee, *et al*. 2006). ANN is a powerful technique for detecting and classifying fault types. Although ANNs are capable of learning the complex nonlinear relationship between process parameters and fault type in manufacturing processes, they cannot explicitly provide a rule to pinpoint the causes of these faults because of the lack of comprehensibility. In order to overcome these shortcomings, various evolutionary algorithms (EA) employed in knowledge discovery in database (KDD) and Data Mining (DM) techniques are used to discover rules in ANNs. For instance, there are currently two GA-based approaches for rule discovery; GA-based rule discovery and extraction algorithm to acquire rules from the manufacturing process, which are used to express the causal relationships between process parameters and product output measures. Moreover, these rules are used to construct an ANN with high performance for monitoring abnormal signals in manufacturing systems. It is crucial to improve the product quality through applying these intelligent approaches to remote monitoring and diagnosis.

Intelligent Manufacturing System

A manufacturing system can be conceptually perceived as being an integration of complex interacting subsystems, organized in such a way as to synergically aspire towards a common set of goals (Merchant, 1984). A manufacturing system is a multi-objective optimization system and usually made up of many subsystems (components) cooperating and coordinating together to function as an integrated system. The IMS was initiated by Professor H. Yoshikawa of the University of Tokyo in 1989. It, as described by him (Mitchell, 1993) is

a system which improves productivity by systematizing the intellectual aspect involved in manufacturing, flexibly integrating the entire range of corporate activities - from order booking through design, production, and marketing - so as to foster the optimum in the relationship between men and intelligent machine.

IMS differs from a conventional manufacturing system - even an advanced one - in its inherent capability to adapt to changes without external intervention (Sousa & Ramos, 1999). It

is an advanced manufacturing system in which production activities from product ordering, designing, and production to marketing are flexibly integrated, and various intelligent manufacturing activities and intelligent equipment are effectively combined. They emphasize the autonomy of the manufacturing unit, and self-organization, learning, and adaptation of the manufacturing system. The attributes of an IMS are as follows.

Self-Organization and Flexibility

The self-organization and flexibility are significant and basic characteristics of the IMS. These attributes can help manufacturing systems organize autonomous functional subsystems to form a running structure in response to new market conditions and work without any outside intervention. In addition, it enables manufacturing system to rapidly adjust their production capacity and functionality.

Human-Machine Integration

Intelligent manufacturing systems are more than just AI-based systems; they have human-machine integration interfaces, that is, hybrid intelligent systems. Machine intelligence based on AI only have defined mechanisms for reasoning, prediction, and decision making. It is not the same as human involving three different types of thought: logical, image, and inspiration. It only has the capacity for logical thought and, at most, part of the ability of image thought, but it never has the ability of inspiration thought. Hybrid intelligent systems characterized by the nature of human-machine integration are the trend of future intelligent manufacturing systems. With coordination between humans and intelligent machines, both can exert their potential on different levels.

Self-Discipline

An IMS composed of many different subsystems can response to dynamically changing environments, thus it has the power to regulate its control strategies and to make production planning dynamic and autonomous. It also makes and modifies its decisions in terms of the information. The attribute of self-discipline can help the entire manufacturing system resist interference, adapt to environments, and have the ability to handle redundancy.

Self-Learning and Self-Maintenance

Self-learning and self-maintenance are two important attributes of IMS. The capability of IMS will need to be continuously improved and promoted through refining its knowledge by self-learning. The IMS has also the ability to self-diagnose during the process of running and recover from the states of malfunction depending on its own capability. Hence, these attributes, self-learning and self-maintenance, give an IMS the power of self-optimization so that it is adaptive to dynamically changing environments.

Intelligent Integration

An IMS is composed of many different subsystems, including management decision, purchase, product design, production planning, manufacturing assembly, quality assurance, marketing, and so on. Although the automatization and intelligentization of each subsystems is the foundation of IMS, the automatization and intellectualization of the entire manufacturing system via effective integration of individual subsystems will be the future objective of development endeavors.

We know from the attributes of IMS including self-organization, self-learning, adaptation and the optimal execution of manufacturing processes,

multi-agent technology with its potential attributes of autonomy, intelligence, pro-activity and reactivity, and self-organization are very suitable for the development of IMS.

SUMMARY

Manufacturing industries are currently facing tough global market competition and challenges in their business environment. Manufacturers must produce a variety of products effectively to meet high customer satisfaction and the expectation of different segments of the market. Intelligent manufacturing is an effective manufacturing paradigm in 21st century manufacturing for this changing environment. Intelligent decision-making and manufacturing process optimization with AI techniques are powerful means to minimize product cost, improve product quality and reduce the lead time in production. MI as a fundament of intelligent manufacturing is the application and extension of AI in manufacturing. MI, as a new discipline of manufacturing engineering, has many scientific foundations and key technologies to be further developed and researched.

REFERENCES

Azuma, R. (1997). A Survey of Augmented Reality. *Presence: Teleoperations and Virtual Environments*, *6*(4), 355–385.

Burdea, G., & Coiffet, P. (2003). *Virtual Reality Technology*. Hoboken, NJ: Wiley-Interscience.

Castillo, O., & Melin, P. (2003). *Soft Computing and Fractal Theory for Intelligent Manufacturing*. New York: Physica-Verlag.

Engelbrecht, A. P. (2007). *Computational Intelligence: An Introduction*. Chichester, UK: J. Wiley.

Koren, Y. (1983). *Computer Control Manufacturing System*. New York: McGraw-Hill.

Kumar, S. (2007). Process Control in CNC Manufacturing for Discrete Components: A STEP-NC Compliant Framework. *Robotics and Computer-integrated Manufacturing*, *23*, 667–676. doi:10.1016/j.rcim.2007.02.015

Kumar, S. R. T. (1995). Intelligent manufacturing Systems: Mapping of AI Tools to Manufacturing. In *Proceedings of 2nd Meeting on Applications of AI Manufacturing Engineering, CIRP, August, Enschede*, (pp. 119-145).

Lee, J., Ni, D. D., Qiu, H., & Liao, H. T. (2006). Intelligent Prognostics Tools and E-Maintenance. *Computers in Industry*, *57*, 476–489. doi:10.1016/j.compind.2006.02.014

Masory, O., & Koren, Y. (1983). Variable Gain Adaptive Control System for Turning. *Journal of Manufacturing Systems*, *2*, 165–173. doi:10.1016/S0278-6125(83)80029-6

Merchant, M. E. (1984). Analysis of Existing Technological Forecasts Pertinent to Utilization of Artificial Intelligence and Pattern Recognition Techniques in Manufacturing Engineering. In 16th CIRP International Seminar on Manufacturing Systems, 14(1), 11-16.

Mirkin, B. G. (2005). *Clustering for Data Mining: A Data Recovery Approach*. Boca Raton, FL: Chapman & Hall.

Mitchell, J. D. (1999). The Intelligent Manufacturing Systems Initiative. In *15th IEEE/CHMT International Electronic Manufacturing Technology Symposium*, 4-6 Oct. (pp. 427-431).

Rao, M., Wang, Q., & Cha, J. Z. (1993). *Integrated Distributed Intelligent System in Manufacturing*. London: Chapman & Hall.

Reti, R., & Kumara, S. R. T. (1997). Intelligent Computing Methods for Manufacturing Systems. *Annals of the CIRP*, *46*(2), 629–652. doi:10.1016/S0007-8506(07)60883-X

Saravanan, R. (2006). *Manufacturing Optimization Through Intelligent Techniques*. Boca Raton, FL: CRC/Taylor & Francis.

Sousa, P., & Ramos, C. (1999). A Distributed Architecture and Negotiation Protocol for Scheduling in Manufacturing Systems. *Computers in Industry*, *38*, 103–113. doi:10.1016/S0166-3615(98)00112-2

Taisch, M., Thoben, K. D., & Monotorio, M. (2007). *Advanced Manufacturing: An ICT and Systems Perspective*. London: Taylor & Francis Group.

Ulsory, A. G., & Koren, Y. (1989). Applications of Adaptive Control to Machine Tool Process Control. *IEEE Control Systems Magazine*, 33–37. doi:10.1109/37.24837

Wang, K. (2005). *Applied Computational Intelligence in Intelligent Manufacturing Systems*. Adelaide, Australia: Advanced Knowledge International.

Wright, P., & Bourne, D. A. (1988). *Manufacturing Intelligence*. Reading, MA: Addison-Wesley.

Wu, C. L., Haboush R. K., Lyrnburner, D. R., Smith, G. H. (1986, November). Closed-Loop Machining Control for Cylindrical Turning. *Modeling, Sensing and Control of Manufacturing Systems,* (pp. 189-204). New York: ASME.

Chapter 2
Knowledge-Based System

INTRODUCTION

Background

Knowledge-Based System (KBS), a branch research area of AI, has been widely used in interpretation, prediction, diagnosis, debugging, design, planning, monitoring, repair, instruction, and control (Stefik et al., 1982) since it emerged in 1960s. KBS has been recognized as a promising paradigm for the next generation manufacturing systems and there is no doubt that the use of KBS in manufacturing will continue to expand, both in areas of application as well as in depth of knowledge. As a result, factories will benefit a lot, such as improved productivity, more stable and increased yields and increased asset utilization, all leading to improved factory performance. Now KBS are finding an increasing number of applications in almost each stage of intelligent manufacturing, including design, process planning and scheduling, production control, diagnosis and etc. Followed by a case study, the overview over all these applications will be discussed in this chapter after the key technologies of KBS are presented, including knowledge representation, knowledge use, knowledge acquisition and evaluation of KBS.

DOI: 10.4018/978-1-60566-864-2.ch002

Basic Conception of KBS

Knowledge-based systems are problem solving systems based on knowledge. Actually, a KBS is an intelligent computer program which is able to solve complex problems in specific areas by imitating the human expert thinking. Thus, besides the large amount of knowledge of human experts, KBS should have the reasoning ability and be able to use this knowledge to solve practical problems, just acting as human experts. For example, a medical KBS can be like a real doctor, diagnose the diseases of patients, determine serious or not, and give appropriate prescriptions and treatment recommendations.

When a KBS is restricted to a very narrow specific domain of expertise and demonstrates expert-level problem solving abilities, it is known as an Expert System (ES). Actually, as two terms, ES and KBS are often interchangeable; there is no real distinction, just subtle differences between them. An ES involves special methods of problem

solving by getting knowledge from human experts, while KBS is more generic and does not emphasize the special process of knowledge acquisition (Li & Qing, 2003). In other words, an ES is a KBS, thus it is often called a Knowledge-based Expert System. In this chapter, the two terms are considered to be interchangeable and to have the same meaning.

Compared to human experts, KBS has its unique advantages. The knowledge of KBS is extracted from the experience of human experts. It is the expertise which is selected and transformed to form a knowledge base, so a fine KBS often has more useful knowledge than a single human expert. Unlike human experts, KBS would never sleep, resign, get sick or retire. It often takes time and normally expensive to train human experts while KBS can be duplicated easily without limitation. In addition, the knowledge of human experts will be lost due to the death of the experts, but the knowledge base of KBS can be updated frequently.

The foundation of KBS dates back to about sixty years ago when Alan Turing presented a influential paper on machine intelligence, 'Computing machinery and intelligence' (Turing, 1950). In this paper he defined the intelligent behavior of a computer as the ability to achieve human-level performance in cognitive tasks and provided a game, the Turing imitation game, to test whether machines could pass a behavior test that requires intelligence. Although even modern computers have failed the Turing test, it provided a basis for the verification and validation of the KBS.

Research on KBS actually began in the mid-1960s. During this period, McCarthy, in his paper 'Programs with common sense', proposed a program called Advice Taker to search for solutions to general problems of the world (McCarthy, 1958). This gave rise to the technologies of knowledge representation and reasoning, as a result of which it is said that the Advice Taker was the first KBS in a true sense. Another ambitious project in this era was the General Problem Solver (GPS) (Newell and Simon, 1961, 1972). However the GPS failed because it is too weak to solve complicated problems in practical applications (Newell and Simon, 1963; Newell, 1969) and consequently kept KBS research in the 1960s in the dark.

Then by the mid-1970s, a group of scientists led by Edward Feigenbaum (Stanford University) gave up the attempt to find several very strong and general problem solvers, narrowed their focus, and developed DENDRAL - in fact, the first successful KBS (Feigenbaum, 1971). This exploration caused a fundamental change in research on KBS - from the extensive exploration of universal laws of human thinking to problems on knowledge, the center of intelligent behavior. Since then, some other KBS have been developed; the most famous one is MYCIN developed by Harvard University for the diagnosis of infectious blood diseases (Shortliffe, 1976).

Now KBS has been complemented by other technologies very well. Artificial Neural Network (ANN), formulated by the late 1960s (Cowan, 1990), can be used for extracting hidden knowledge in data base to form rules for KBS (Medsker and Leibowitz, 1994; Zahedi, 1993) and correcting rules in the knowledge base of KBS (Omilin and Giles, 1996). Fuzzy logic, introduced by Professor Lotfi Zadeh (Zadeh, 1965), can be used for improving computational power and cognitive modeling of KBS (Cox, 1999; Turban and Aronson, 2000). As a result, the KBS has improved a lot in adaptability, fault tolerance, and speed. Thus a large number of KBS have been successfully developed and applied into various fields. Currently, KBS are mainly used in the following areas: explanation, diagnosis, monitoring, forecasting, planning and design.

Explanation

KBS for explanation analyze and measure the observed data, then determine its meaning. The data are often from various sensors and measuring instruments; for example, DENDRAL analyzes the

data observed by mass spectrograph to determine the molecular structure of the tested compounds. The molecular structure is the interpretation of the metrical data. In order to find interpretations which are correct and compatible with the data, the data analysis must be comprehensive; therefore, the analysis system should inspect all possible explanations systematically and must have adequate evidence to select the candidate answers. It is always difficult to get the correct explanation, since the metrical data often contain errors and are sometimes incorrect, so the metrical data must be handled properly.

Diagnosis

KBS for diagnosis determines the possible roots of failure according to the failure phenomena, such as people being sick, machinery breakdown and so on. To make a correct diagnosis, KBS must have the knowledge about the structure, function and behavior of the things being diagnosed. What is more, heuristic knowledge on the association between the roots and the phenomena can accelerate the diagnostic process. The difficulty of the diagnosis task lies in: multiple faults (some faults are covered by other faults); intermittent faults (this requires enhancement of the diagnostic tools); sensors and measuring instruments may have their own faults; manual data measurement may be costly or dangerous (this requires choosing the right type of data to be tested); diagnosis knowledge is inaccurate or incomplete (such as the mechanism of the human diseases).

Monitoring

KBS for monitoring continuously observe and explain the data which reflect the status of a system and equipment. When an exception occurs, KBS should issue a warning, propose a solution and give out the reason. Monitoring is an advanced stage of automatic control, so the system will automatically adjust the parameters to remove anomalies and maintain the system in good running condition. The main basis of monitoring is to continuously compare the data which reflect the present status of the system with those which are observed when the system is under normal operation. Real-time diagnosis of exceptional causes and avoiding false alarms are the key to improving the performance of monitoring.

Forecasting

KBS for forecasting predict the future status and behavior from the current situation and development trend, such as forecasting the influences after economic policy changes, weather forecasts, traffic forecasts, harvest forecasts and military forecasts. Forecasting often infers chronological order, and must deal with things that change with time. The characteristics of forecasting are to carry out comprehensive analysis with incomplete information (forecasting with complete information does not need KBS, such as predicting where the Jupiter will be two years later in its orbit). Moreover, forecasting must consider a variety of possible scenarios and make use of assumption reasoning.

Planning

For planning, KBS aim at establishing programs for action which would meet a variety of constraints, including production planning and scheduling, project planning, military operations planning, robot behavior planning and so on. Planning often involves optimizing the use of resources issues (time, manpower and material resources), as strict constraints could lead to over-planning failure, which then requires minor relaxation of constraints. When the planning problem is large and complex, the most important planning objective should be given priority, and the interaction that may arise between multiple objectives should be weighted properly.

Design

Architectural design, especially electronic circuit design, is a dynamic field suitable for KBS. Design and planning are similar in many ways, but design is more complex. It not only needs to meet a variety of restrictive conditions, but also consider issues about the spatial relationship of parts and lines, such as distance, shape, appearance and so on, which would consume a large amount of computing resources. A better method for spatial relationship inference does not yet exist, while how to satisfy the restrictions and the special requests of design is yet another problem.

When implementing these six kinds of tasks, the difficulties that often arise are: the large-scale solution space (explanation, planning and design); requests for exploratory reasoning (because of lack of time, facts and knowledge in diagnosis, planning and design then have to make use of assumptions); time-varying data (monitoring and diagnosis); data with noise (often the attached sensor noise). In order to overcome these difficulties, appropriate inference methods should be designed first, after which the architecture of the KBS can be determined.

Basic Structure of KBS

The basic structure of KBS is shown in Figure 1, consisting of six basic components: knowledge base, inference engine, database, explanation facilities, knowledge acquisition mechanism and user interface.

Knowledge base is the most important part of the KBS. It contains the specific domain knowledge useful for problem solving. The updates of knowledge in knowledge base can improve the performance of KBS.

The inference engine has the capacity for reasoning. According to the input problems and the data which describe the initial state of problem solving, the inference engine carries out reasoning with the knowledge in the knowledge base, and ultimately outputs the answer. In short, it applies domain knowledge to draw conclusions from known facts.

The database is designed to store raw data that needed in the reasoning process, intermediate results and final conclusions, often as a temporary memory.

The explanation facilities enable the KBS to review its reasoning process, explain and justify its analysis and conclusions, which would make users to understand how a particular conclusion is reached.

The user interface is not the main part of the KBS, but a user-friendly interface is very important for the promotion and implementation of the KBS. At present, a visual user interface has been widely used in KBS.

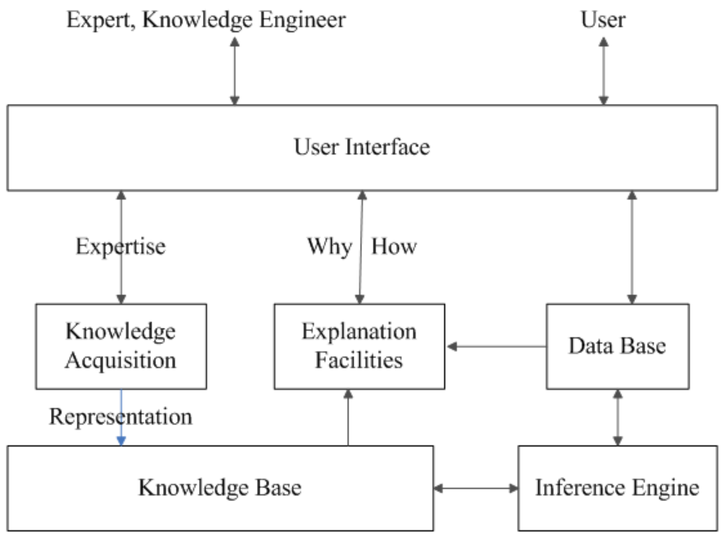
Knowledge representation is the key technology in the process of building a KBS. It acquires human expertise and transforms them into a data presentation that can be understood by a machine system. This technology will be detailed in the next section about knowledge engineering composed of knowledge representation, knowledge use and knowledge acquisition and management.

THE PROCESS OF BUILDING KBS-KNOWLEDGE ENGINEERING

Introduction

In 1977, at the fifth international conference on Artificial Intelligence, Feigenbaum proposed the concept of 'knowledge engineering'. Knowledge engineering mainly studies the relevant technologies that are needed in implementing KBS, and the most important element in KBS is knowledge processing, including knowledge representation, knowledge applications, knowledge acquisition, and the tools and technologies to realize knowledge engineering. The process of building intelligent KBS is called knowledge engineering.

Figure 1. Basic structure of KBS



In the first ten years after AI being created, people focused on the process of problem solving and inference. Through experiments and research, Feigenbaum proved that the primary means of realizing intelligent behavior lies in knowledge, especially knowledge in specified areas in most practical cases. He first proposed the concepts of expert system and knowledge engineering in 1977, which became a milestone in knowledge operations.

In the past 20 years, the research on knowledge engineering has developed significantly. While the connotation of Artificial Intelligence still remains unchanged, research focuses on knowledge engineering have changed greatly. (1) Major research content has been translated from knowledge inference to knowledge representation. (2) The major study objective has been translated from knowledge in a specified domain to common sense, then to engineering knowledge. (3) The processing objectives have changed from standardized, comparatively easily treatable knowledge to non-standardized and relatively refractory knowledge. (4) The scale and method of processing has changed from closed to open, and from small to large scale projections that can deal with a massive expansion of knowledge. (5) The study has changed from knowledge engineering to knowledge science.

Regarding knowledge engineering, different experts and scholars have their own understanding and expositions. Professor Feigenbaum (Stanford University) considers KBS as a computer program with knowledge and ability, which can act just as the human experts and is also known as knowledge engineering. Professor Bynum (Stanford University) holds that knowledge engineering completely changes the definition and use of knowledge, and regards that knowledge engineering is a branch of information technology, often associated with Artificial Intelligence. Professor Guus Schreiber (Free University of Amsterdam, the Netherlands) views knowledge engineering as a methodology used as a tool, and KBS as a key product equivalent to knowledge management. Professor Zhongzhi Shi (The CAS Institute of Computing Technology of China) believes knowledge engineering is a discipline dealing with knowledge information and providing technologies to develop intelligent systems and is the result of interdisciplinary development among Artificial Intelligence, database technology, mathematical logic, cognitive science,

psychology, and others. Academician Rujin Lu (The CAS Institute of Mathematics in China) thinks that, for more than 20 years, knowledge engineering has mainly been an experimental science, and a large number of theoretical issues on knowledge processing have not yet been resolved. The study of knowledge should be a science with solid theoretical foundation, and the concept of knowledge engineering should be transferred into knowledge science.

From the viewpoints above, it is easy to see that the understanding of most well-known experts and scholars on knowledge engineering is limited to expert systems, artificial intelligence, knowledge management, and experimental science. It is basically engineering on knowledge rather than knowledge on engineering. There was a time when some people considered expert systems as knowledge engineering, which reflects a certain limitation of understanding. In fact, looking at development in recent years, the connotation of knowledge engineering has moved far beyond the scope of expert systems, and has gradually extended and deviated from the early definition of knowledge engineering.

The study contents of knowledge engineering include: relative knowledge emergence within a complex product life cycle (new knowledge acquisition); representation (formalization); organization (systematism); sharing (knowledge transfer); search (getting access to the knowledge); use (materialization), and update (outdated knowledge elimination). Each aspect is explained in more detail as follows:

- **The emergence of knowledge:** new knowledge acquisition, including knowledge mining (also called 'knowledge mining', 'knowledge extraction', 'knowledge discovery'). It stresses the discovery and collection of individual, non-materialized knowledge which is outside the existing management system.
- **The representation of knowledge:** which includes how to represent knowledge in best possible form to facilitate efficient understanding.
- **The organization of knowledge:** which builds associated mutual relations among knowledge and realizes the effective organization of knowledge.
- **The sharing of knowledge:** which realizes the high degree of sharing and rapid transmission of knowledge by means of the network, and takes the knowledge base as the core.
- **The search of knowledge:** efficiently querying and retrieving relevant knowledge based on ontology relations, then realizing the access to the knowledge.
- **The use of knowledge:** making full use of knowledge in the product life cycle, making the knowledge fully available, enabling the intellectual assets to be delivered into physical assets, and allowing the knowledge o create value for enterprises.
- **The update of knowledge:** regularly eliminating the outdated knowledge and maintaining the effectiveness and availability of the knowledge base.

In the next section, knowledge representation, knowledge acquisition and management, knowledge use including inference techniques and conflict resolution strategies in knowledge engineering are introduced.

Knowledge Representation

What is Knowledge?

Knowledge is the summation of understanding and experience accumulated by humans. It is a theoretical or practical understanding of a subject or domain. In an expert system, the knowledge in the knowledge base is only a part of the knowledge that relates to the problem field. There is no

strict definition of what knowledge is, but several representative definitions are shown as below.

1. **Feigenbaum:** knowledge is information that has been cut, shaped, explained, selected and converted.
2. **Bernstein:** knowledge is composed of descriptions, relations and processes of a specific domain.
3. **Hayes-Roth:** knowledge = fact + belief + heuristic.

The knowledge base of expert systems often contains different types of knowledge, such as knowledge about objectives, knowledge about processes, commonsense knowledge and knowledge on goals, motivations, reasons and time, therefore expert systems are also known as KBS. From the role it plays in the problem-solving process, knowledge in the knowledge base can be classified into several types: descriptive knowledge, problem solving knowledge, control knowledge and metaknowledge. Descriptive knowledge illustrates facts, actions and events which are related to the objectives. It describes not only the objectives but also the scope of the objectives and the casual relationships between the incidents. Problem-solving knowledge is the knowledge of a special domain, and it is the key for expert systems to reach a level of performance by means of solving problems comparable to those of a human expert in some specialized application domain. Control knowledge mainly describes the strategies used to solve problems and controls the process of problem-solving. Metaknowledge is the knowledge about knowledge, and it plays an important role in knowledge management, knowledge interpretation, knowledge utilization and knowledge learning.

Representation is a group of agreements made to describe the word, and is the symbolization process of the knowledge.

Knowledge representation encodes the facts, relations and processes about the word into a suitable data structure. The main challenge for knowledge representation methods is to design all kinds of structures; it studies the relationships between representation and control, between representation and inference, relationships between knowledge representation and other scientific areas. At present, knowledge representation is one of the most active research issues in Artificial Intelligence, aiming at making decisions and conclusions through effective knowledge representation (McCarthy and Hayes, 1969).

On knowledge representation, KBS have the following basic requirements:

1. **Strong expressive ability.** KBS require knowledge representation to express the knowledge that is needed to solve the problem correctly and effectively.
2. **Easy to understand.** The knowledge represented should be easy to read and understand, to make the program dealing with the knowledge relatively simple.
3. **Easy to access, modify and expand.** The knowledge in the knowledge base should be used, modified and expanded easily and effectively. The development of a KBS is a process of constantly modifying, expanding, and improving the knowledge in the knowledge base.
4. **Manageable.** The knowledge base should be capable of being easily checked for its accuracy and consistency; also, the knowledge in the system should be debugged easily.

Research on knowledge representation is still under development. The most important approach to test the pros and cons of the knowledge representation methods is their effectiveness in practical applications.

Knowledge representation describes the various data structures used to represent the experts' knowledge. The approaches for knowledge representation are various, with their own advantages and disadvantages. An approach suitable for one

type of problem may not be appropriate for another kind of problem. The most suitable approach should be selected according to the processing objectives.

Major approaches to knowledge representation include logic representations, semantic networks, procedural representations, logic programming formalisms, frames, production rules and predicate calculus (Barr and Feigenbaum, 1981; Cercone, 1987). However, production rules and frames are the most frequently used in commercial expert system shells.

Production Rule Representation

The human mind is very complex and cannot be represented by any algorithms. Nevertheless, most experts can express their knowledge in production rules which can be used for problem-solving. Consider a simple example: if there are three countries, China, France and Ireland, and three capital cities, Paris, China and Dublin, and you are asked to match each country with its capital city, then how would you do this?

It is common sense that we know that China's capital city is Beijing, France's capital city is Paris and Dublin is Ireland's capital city. Thus you have three simple rules. Your knowledge is formulated as the following simple statements:

```
IF the 'country' is China
THEN the 'capital city' is Beijing

IF the 'country' is Ireland
THEN the 'capital city' is Dublin

IF the 'country' is France
THEN the 'capital city' is Paris
```

The statements represented above in the IF-THEN form are known as production rules or just rules. Production rule is the most commonly used type of knowledge representation and can represent relations, recommendations, directives, strategies and heuristics (Durkin, 1994); it was first developed by Newell and Simon (1972), and expert systems using this type of knowledge representation are called rule-based expert systems.

All production rules consist of two parts: the IF part that states the conditions that must be presented for the production rule to be applicable, is called the condition (antecedent or condition); and the THEN part that represents the appropriate action to take, is called the action (conclusion or consequent). During the execution, a production rule whose condition part is satisfied can be executed by the inference engine. The resulting action will then usually generate new conditions or facts to be added to the knowledge base. Compared with other approaches to knowledge representation, production rules are relatively easy to create and understand.

The basic syntax of a rule is:

```
IF    <condition>
THEN  <action>
```

In general, a rule can have multiple conditions which are usually joined by the keywords AND (conjunction), OR (disjunction) or a combination of both. However, it is better to avoid mixing conjunctions and disjunctions in the same rule.

```
IF    <condition 1>
AND   <condition 2>
      ...
      …
      …
AND   <condition n>
THEN  <action>
IF    <condition 1>
OR    <condition 2>
      …..
      …..
      …..
OR    <condition n>
THEN  <action>
```

A rule may also have multiple actions:

```
IF    <condition>
THEN  <action 1>
  …   <action 2>
  ….
  ….
  …. <action n>
```

For example, a rule in a KBS for Computer-Aided Process Planning (CAPP) is as below.

IF the outer surface is cylindrical or conical

AND machining accuracy level IT values equal to or greater than 6

AND surface roughness is between 6.3 and 40

THEN turning

In conclusion, KBS that use rules as a knowledge representation technique have the following advantages:

1. Uniform structure. All production rules have the syntax IF-THEN structure, which enables them to be self-documented.
2. It is easy to add, delete and update information. A knowledge engineer can simply add rules to the knowledge base without intervening in the control structure.
3. Naturalness, which is to say, natural knowledge representation. The knowledge can be represented quite naturally in IF-THEN production rules, which are very close to the expressing methods of cause and effect, visual, natural and easy for reasoning.
4. It is conducive to heuristic knowledge and allows guiding the process of interpretation by using domain knowledge.

The advantages of rule-based expert systems make production rules highly desirable for knowledge representation in real-world problems, but they also have some shortcomings:

1. **Ineffective search strategy.** During each cycle, the inference engine will apply an exhaustive search through all the production rules. Therefore, rule-based systems with a large set of rules can be very slow and ineffective, and large rule-based systems can be unsuitable for problem-solving in real-time.
2. **Opaque relations between rules.** Although the individual production rule tends to be relatively simple and powerful to represent knowledge with relationships, they are not good at expressing structural relationships between items and their logical interactions within the large set of rules. This is due to the inadequate hierarchical knowledge representation in the rule-based expert systems.
3. **Inability to learn.** In general, expert systems still need a knowledge engineer to be responsible for revising and maintaining them, as they do not have the ability to learn from experience. They do not know how to 'break the rules', or to modify their knowledge bases, adjust existing rules or add new ones.

Frame Representation

A frame is a partition of knowledge which consists of a collection of slots to describe a particular object or concept. It was first proposed by a well-known AI scholar Marvin Minsky in the 1970s (Minsky, 1975). A frame can combine all necessary knowledge about a particular object or concept, including its various attributes, states and relationships with others. It is an improved hierarchical knowledge representation method with uniform structure.

A frame can be considered as a data structure with typical knowledge to describe an object. Each frame has its own name and a set of attributes or slots associated with it. For example, name, weight, height, and age are slots in the frame "person". Engine type, horsepower, transmission type and seating capacity are slots in the frame "car".

Table 1. A frame for lathe

Lathe			
	Type	The number of shafts	Uniaxial, Multiaxial, etc
		The degree of automation	Automatic, semi-automatic, manual, etc
		The largest swing diameter	320mm, 400mm, 600mm, etc
	Structure	Headstock	Frame for headstock
		Electrical control cabinet	Frame for Electrical control cabinet
		Lathe bed	Frame for lathe bed
		Tool carrier	Frame for tool carrier
		Tailstock	Frame for tailstock
		…	

The content of the frame is a list of slots that define relationships to other frames which play significant roles in a particular situation. Attached to a frame or a slot are different kinds of information such as default values, constraint specifications, procedures and so on. For example, a frame for a "teacher" has slots for its name, gender, age, salary, address, telephone, etc., to describe a particular teacher. A 'copy' of this teacher frame can be created, and the slots can be filled in with the information about the particular teacher being described. For example, a frame for a particular teacher named Missy is shown as below.

```
Frame:        teacher
Name:         Misshy
Gender:       Female
Age:          38
Title:        professor
Salary:       4000RMB
Units:        Wuhan University
Address:      Wuhan
Telephone:    027-50857185
```

The slot can also be divided further—take the slot 'salary' as an example:

```
<salary>
   Basic pay:
   Service wage:
   Allowance:
   Moth bonus:
   ……….
```

Therefore, a frame resembles related 'record', and slots are used to store values. A slot may contain a default value or a pointer to another frame, a set of rules or procedures by which the slot value is obtained. The general form of a frame is described as below.

```
Frame name
Slot 1  Side 11(value 111, value 112)
        Side 12(value 121, value 122)
        ……
Slot 2  Side 21(value 211, value 212)
        Side 22(value 221, value 222)
        ……
Slot 3  Side 31(value 311, value 312)
        Side 32(value 321, value 322)
        ……
……
```

As examples, Table 1. is a frame for a lathe and Table 2. is a frame for a headstock.

In fact, the frame can be divided into two types, class-frame and instance-frame. Class-frame may refer to a group of similar objects while instance-frame infers a particular object. Consider a simple example where a frame com-

Table 2. A frame for headstock

Headstock			
	Spindle parts	Spindle	Shaft end, shaft journal, shaft lever, etc
		Bearing	Front bearing, back bearing, intermediate bearing, etc
		Oil seal	Oil slinger, asphaltic felt, spacer, etc
	Box parts	Enclosures	Cover, box, control panel, rating plate, etc
		Coupling parts	Hinge, screw nut, etc
	Shaft parts	Drive shaft	The first shaft, the second shaft
		Gear	Change gear, sliding gear, idler gear, etc
	Control unit	Handle	Change lever 1, change lever 2, etc
		Button	Start button, stop button, etc
	Lubrication	Pump	Gear oil pump, needle roller bearing, etc
		Pump motor	Stator, rotor, binding post, cooling fan, etc
		Oil pipe	Oil inlet, pipe joint, etc

puter can be a class-frame, while computer IBM Aptiva S35 is an instance-frame. We will use the instance-frame when referring to a particular object, and class-frame when referring to a group of similar objects.

A class-frame describes a group of similar objects which have common attributes. For example, desk, person, car and computer all are considered as class-frames.

Each frame in a frame-based system should know what class it belongs to. That is to say, the frame's class is an implicit attribute or characteristic of the frame. For example, instances in Table 2 identify their class in the slot frame.

The knowledge base which uses frames for knowledge representation consists of a set of structural frames. Its main advantages are:

1. It has strong ability to represent knowledge.
2. Rich hierarchical structures provide an effective means of knowledge organization.
3. The reasoning is flexible.
4. The reasoning is driven by expectations.

Knowledge Use

Human experts have the ability to resolve complex practical problems because they have a large quantity of knowledge in a special domain, also they have the ability to reason and generate new knowledge.

When a knowledge-base system has acquired the knowledge, the system should consider how to search and use the knowledge in reasoning, which indicates that reasoning is related to the knowledge. In expert systems, it is the inference engine that carries out the reasoning with knowledge. The methods for the inference engine to schedule and use knowledge are called the control strategy. The control strategy in a system determines the speed of solving problems, which can affect the success or failure of the system. Inference techniques and confliction resolution techniques are the most important parts of the control strategy.

Inference Techniques

Inference is the process of getting new information (conclusion) from the known information (the premise) based on certain principles. The

inference engine is the backbone of the KBS. It interprets knowledge in the knowledge base and makes reasoning among knowledge rules to reach conclusions for the domain (Barr and Feigenbaum, 1981; Rowe, 1988; Winston, 1984).

Taking the rule-based expert system as an example, the domain knowledge is represented in a collection of production rules in "IF-THEN" form, and the data about the current situation is represented by a set of facts. The inference engine is responsible for comparing each rule in the knowledge base with the fact in the database. When the IF part (condition) of a rule matches the fact, the rule will be fired or executed. The matching of the rule IF parts to the facts produce inference chains, which indicate how an expert system applies the rules to reach a conclusion.

According to the inference direction, the inference can be classified into forward inference, backward inference, forward and backward mixed inference, and imprecise inference.

Forward Inference

Forward inference technology is also known as data-driven reasoning, bottom-up reasoning, forward reasoning, forward chaining and antecedent reasoning. The rules are run forward, deducting all logical conclusions from a basic set of input data. As a data-driven reasoning technology, forward inference starts with the input data and proceeds forward with that data until it achieves the goal.

The general process of forward inference is that firstly, the user inputs the original information or data, according to which the inference engine searches for rules in the knowledge base to find a rule of which the IF condition would match the input data. If the rule is found, the conclusion part of the rule can be seen as the intermediate result. Then the inference engine will search in the knowledge base for rules matching the intermediate result, until the final conclusion is reached. Its algorithm includes the following steps:

- **Step 1:** user inputs the relative information into the data base
- **Step 2:** check if the solution has existed, if so, the end; if not go to the next stop
- **Step 3:** based on the current state, choose proper knowledge in the knowledge base to reason and generate new information, and store the new information into the data base
- **Step 4:** go to step 2 until the goal user set is reached

The algorithm above initially appears to be simple, but in fact it is very complex to implement. For example, selecting appropriate knowledge in the knowledge base according to the current state involves the kind of methods you will use to search; judging whether the knowledge is available or not is even more complex, as we rarely use direct match but consider the form of the problems and the condition to use the knowledge; if the proper rule is reached more than once, the problem arises as which to execute first.

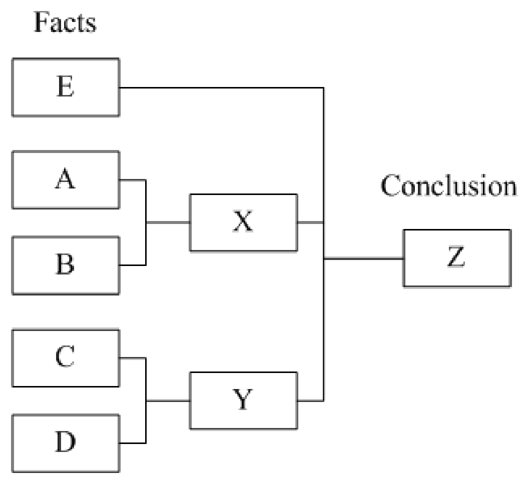
In order to illustrate the forward inference technique, let us consider a simple example.

Suppose the data base initially includes facts A, B, C, D, and E. and the knowledge base contains the following three rules:

```
Rule 1:     IF      E is true
            AND     X is true
            AND     Y is true
            THEN    Z is true
Rule 2:     IF      A is true
            AND     B is true
            THEN    X is true
Rule 3:     IF      C is true
            AND     D is true
            THEN    Y is true
```

The inference process shown in Figure 2 indicates how to infer the fact Z by applying the rules

Figure 2. Example of forward inference process



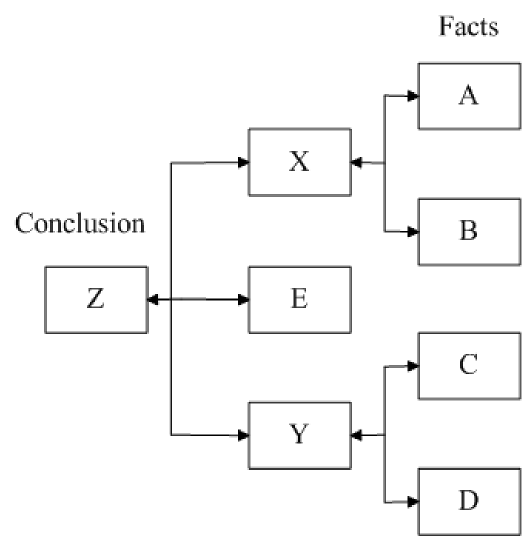
in the knowledge base. First rule 2 is executed to deduce new fact X from the initially given facts A and B. Rule 3 is then fired to deduce new fact Y from given facts C and D. Finally, rule 1 applies initially known fact E and just obtained facts X and Y to reach the conclusion Z.

The main advantage of positive reasoning is that it can take full advantage of the information provided by the user, and can quickly respond to information input by the user. Its drawback is that it is the inference of a very weak purpose, and is intended to be blind reasoning. It gathers all the information and then infers from the knowledge whatever can be inferred, without a specific goal. There would therefore be many rules which have nothing to do with the ultimate goal being executed, and furthermore, it would lower the efficiency of the system. In a word, if a specific goal or fact is to be achieved, forward inference would not be efficient enough. Under such circumstances, backward inference may be more appropriate.

Backward Inference

Backward inference is also known as goal-driven, top-down, backward chaining and consequent reasoning. It supposes a conclusion or goal first,

Figure 3. Example of backward inference process



and then proceeds with this goal, and finds further evidence to support the conclusion. The inference steps are:

- **Step 1:** a hypothetical solution or a goal is proposed
- **Step 2:** check the database to establish whether the hypothesis exists or not; if it does, the end; if not, go to the next step
- **Step 3:** find out whether the hypothesis is original data; if it is, tend to the user; if not, go to the next step
- **Step 4:** search the knowledge base to find rules that might have the desired solution. The THEN (action) parts of such rules should match the goal we have set
- **Step 5:** set the IF (condition) parts of the rules we find as our new goal, a sub-goal, go to step 2 and repeat the process until no rules are found in the knowledge base to prove the current sub-goal.

How backward inference works is shown in Figure 3. Take the forward inference as an example and rewrite the rules in the following form:

```
Rule 1: E & X & Y → Z
Rule 2: A & B → X
Rule 3: C & D → Y
```

Suppose the data base initially includes facts A, B, C, D, and E, and we set fact Z as our goal.

- **Step 1:** the inference engine tries to infer the fact Z by searching the knowledge base to find the rule which has the goal, the fact Z, in its THEN (action) part. In our case the inference engine will find the Rule 1: E & X & Y → Z. The IF part of the rule 1 includes facts E, X, and Y, these facts therefore should be determined in order to infer the fact Z.
- **Step 2:** the inference engine sets the fact Y as the new goal and tries to establish it. It first checks the database, when it fails, searches the knowledge base again for the rule that infers the fact Y. The inference engine will find Rule 3: C & D → Y. the IF part of the rule 3 includes facts C and D, so these facts have to be established.
- **Step 3:** the inference engine sets the fact X as the sub-goal. It checks the database, but fact X is not there. Then it searches the knowledge base for the rule with fact X in its THEN part. Finally the inference engine will get the Rule 2: A & B → X. The facts A and B should be determined.
- **Step 4:** the inference engine finds facts A and B in the database, Rule 2: A & B → X is executed and the new fact X is inferred and added to the database.
- **Step 5:** the inference engine finds facts C and D in the database, Rule 3: C & D → Y is executed and the new fact Y is inferred and added to the database.
- **Step 6:** the inference engine returns to the goal fact Z and once again tries to execute Rule 1: E & X & Y → Z. Now the IF part of the rule 1 all matches the facts in database, rule 1 is executed and the original goal Z is finally determined.

At present, there are two approaches to propose the hypothesis or the goal. The first one is proposed by the user, but the degree of automation is very poor; the other is to design a procedure so that the system can propose assumptions automatically according to the current status.

The main advantage of the backward inference is that rules that have nothing to do with the ultimate goal will not be executed. It is highly object-oriented and the only data used is the data needed to support the direct line of reasoning. Its shortcoming lies in its blindness in the choice of an initial goal.

Forward and Backward Mixed Inference

The main disadvantage of forward inference is its weak reasoning purpose, which leads to wasteful work that has nothing to do with the problem solving. The main shortcoming of backward inference is the blindness in its initial goal; some targets it establishes might be false and invalid. With large solution space, this poor situation may be more serious. Forward and Backward Mixed Inference is proposed as an effective reasoning method, which combines the advantages of both forward and backward inferences..

Based on the original information, forward and backward mixed inference first obtains an uncertain conclusion by making use of forward inference, and then attempts to find evidence to prove the conclusion inferred by forward inference by making use of backward inference, repeating this process until the final conclusion is determined.

In mixed inference, backward inference can also be executed before forward inference. In such a situation, the goal is firstly set, and then the backward inference is executed to get some information about the problem, thereafter the forward inference is acted upon by using the

just-obtained information to get the solution to the problem.

Imprecise Inference

In real life, people often work with inaccurate or incomplete knowledge and information. Expert systems work in similar situations as human experts, so it is inevitable that KBS will deal with imprecise knowledge. The inference based on imprecise knowledge is called imprecise inference.

Obviously, imprecise inference seeks to get a reasonable or nearly reasonable conclusion by inferring with the uncertain information or uncertain relationship in real world. Nowadays, most KBS have the function of imprecise inference.

Generally speaking, there are two types of uncertainty: the evidence (assertion) uncertainty and the knowledge uncertainty.

Evidence uncertainty occurs mainly because the evidence or assertion itself may be wrong and the relevance is inaccurate or incomplete. A great deal of evidences come from observation, but when people observe, they often view the facts with uncertainty.

Knowledge uncertainty lies in the immaturity of the knowledge itself, such as experience knowledge in a specialized field, and the neglect of the knowledge's secondary factors.

Usually, an imprecise inference model includes: (1) Description of evidence or assertion with uncertainty; (2) Description of knowledge with uncertainty; (3) Update algorithm with uncertainty.

At present, imprecise inference methods which are widely used in KBS include: (1) certainty factors theory; (2) Bayesian reasoning; (3) fuzzy reasoning; (4) evidential reasoning.

Conflict Resolution Strategies

Conflicts often occur during the reasoning process and in order to get a correct answer, the conflicts need to be resolved. Consider a simple example: in a water PH value case, it's known that if the PH value is less than 0.6, the water is acidic; if the PH value is between 0.6 and 0.8, the water is normal; if the PH value is greater than 0.8, the water is alkaline. Thus it is very simple to decide whether the water is acidic, normal or alkaline according to its PH value and the knowledge can be formulated as rule 1, rule 2 and rule 3 below.

```
Rule 1: IF    'PH < 0.6'
        THEN  the 'PH value of
              water' is acidic
Rule 2: IF    '0.6<PH < 0.8'
        THEN  the 'PH value of
              water' nearly normal
Rule 3: IF    'PH > 0.8'
        THEN  the 'PH value of
              water' is alkaline
Rule 4: IF    'PH < 0.6'
        THEN  the 'PH value of
              water' is alkaline
```

For the given four rules above, what result we will get if the condition is 'PH<0.6'?

During the reasoning process, an inference engine compares the IF (condition) parts of the rules with facts available in the database, and when the condition is satisfied, the rule is executed. So when the fact 'PH<0.6' is in the database, rule 1 and rule 4 will be selected at the same time, for they have the same IF part. When the condition part is satisfied, both of them can be set to fire. In reality, the inference engine can execute only one rule at a time, so the inference engine has the task of determining which one is to be fired. The strategy for choosing a rational rule to fire when a number of rules are available is called conflict resolution.

So, how can we determine which rule will be fired? One commonly used strategy is to establish a goal and stop the rule execution when the goal has obtained a value. Taking the water PH value case as an example, the goal or target is to establish a value for the object 'PH value of

water'. Once the object 'PH value of water' attains a value, the rule execution will be stopped. Therefore, when the PH value is less than 0.6, rule 1 is to be found first, fired, and the 'PH value of water' obtains the value 'acidic'. The rule execution is then stopped, so rule 4 will not be fired. In the given example, the right conclusion can be achieved, but if the order of the rules is changed, rule 4 might be executed while rule 1 is stopped, and the wrong conclusion is then drawn. In this method for conflict resolution, the rule order in the knowledge base is very important.

In general mathematical problem-solving, a conflict resolution strategy called depth-first or breadth-first strategy is often used. Its main idea is to fire one of the available rules randomly and if it leads to the wrong action or conclusion, go back to the other rules and try another one. But in the problem-solving environments of expert systems, such a strategy is often inefficient and sometimes even intolerable. On the one hand, with the increasing complexity of real problems, the work of searching and selecting rules will be very large in an attempt to find out every possible solutions to the problem, it is very difficult to get a conclusion or solution in a limited period of time. On the other hand, some practical issues, such as emergent patient and production process, is real-time monitoring, they demand a very high speed response and cannot allow to give out a solution after testing all the possible solutions.

Besides the method above, there are other conflict solving strategies, such as selecting the right with the highest priority, choosing rules arbitrarily, executing the rule that uses the data most recently entered in the knowledge base and fire the most specific rule and so on.

Conflict resolution strategy is a basic control strategy and often involves other control strategies.

Knowledge Acquisition and Management

Knowledge acquisition is the bottleneck in the development of KBS. The core of an expert system is its knowledge, and this is why expert systems are often called Knowledge-based systems. The current study is not limited to the development of specific KBS. With the development of tools such as knowledge representation language, combination development tools and skeleton systems, study on KBS has entered a new area. It now focuses on the construction of the development environment, which provides a ready-made alternative for knowledge representation and knowledge utilization. However, knowledge acquisition is still an outstanding problem in KBS development owing to lack of efficient development tools. Knowledge acquisition is the first step in the construction of KBS, and it is also the most difficult and one of the most important parts. In this section, the tasks, methods and steps of knowledge acquisition are discussed.

The Task of Knowledge Acquisition

Knowledge acquisition is required to extract special knowledge for problem solving from knowledge sources, and convert the knowledge into a form that can be executed by computer. It is not a one-way knowledge accumulation. On the contrary, mutual influences exist between knowledge acquisition and system construction. Knowledge sources are the object of knowledge acquisition, and are varied. They could be books, literatures, domain experts and so on. The most difficult type to obtain is experience knowledge from domain experts. The complexity of the knowledge sources determines the complexity of the knowledge acquisition.

Whether a KBS is good or not depends on the quality and quantity of its knowledge in the

knowledge base. That is, in order to improve the reliability, effectiveness and availability of the system, KBS must have detailed and refined knowledge.

Refining knowledge is not an easy task, as knowledge not only comes from books and literature, but also experience accumulated in a long-term practice that is known as heuristic knowledge. Generally, heuristic knowledge is very difficult to express, but plays an extremely important role in practical applications. For example, a graduate majoring in computer science acquires knowledge about computers and electronics in the university, but he would not be able to resolve a computer hardware failure. That is because he does not know how to apply the knowledge he has learned in practical use, and he needs to gather and accumulate experience constantly in long-term practice.

Refining knowledge is one part in the process of knowledge acquisition, including the understanding, extraction, and organization of the existing knowledge, and obtaining new knowledge from the existing knowledge and facts. During the process of acquiring new knowledge, the following rules are normally included:

1. **Accuracy:** the knowledge acquired should accurately represent the experiential thinking methods of the domain experts
2. **Reliability:** the knowledge should be recognized and understood by most domain experts, and be able to be tested in practice.
3. **Integrity:** the knowledge should maintain the consistency and integrity of the knowledge collection.
4. **Refining:** making sure the knowledge collection has no redundancy.
5. The methods of knowledge acquisition and its classification.

The complexity of the knowledge sources determines the diversity of the knowledge acquisition methods. Knowledge acquisition has two important components; the one is to obtain information from the external environment, the other to systemize the information obtained in the internal system. According to the different reasoning abilities of the learning systems, there are various forms of knowledge acquisition, and the forms of the acquired information also vary. Knowledge acquisition methods can be divided into the following categories according to their capacity.

1. Knowledge acquisition methods with no reasoning ability (that is, artificial acquisition methods).
2. Knowledge acquisition methods by using editing tools (that is, semi-automatic knowledge acquisition methods).
3. Knowledge acquisition methods with reasoning capacity (that is, automatic knowledge acquisition methods). This can also be divided into interpret-methods and inductive-methods.
4. Ultra-autonomous knowledge acquisition methods.

In the following sections, three methods of knowledge acquisition are briefly introduced.

Knowledge Acquisition Methods with No Reasoning Ability

In terms of acquiring information, these methods do not need reasoning functions for learning. The information obtained by these methods is all data or programs, which can be used directly to solve problems. The method acquires knowledge through knowledge engineers. Knowledge engineers and experts work together in the same team to extract and refine knowledge, which is the most widely used method for knowledge acquisition. The knowledge engineering team can be composed of an expert and a knowledge engineer, or several experts and a number of knowledge engineers. In the team, the experts provide expertise and

experience to the knowledge engineers, and the knowledge engineers are responsible for conceptualizing and formalizing the knowledge provided by the experts. This method aims at providing a user interface that is friendlier than the physical expression in machines. The principle of systems in the process of receiving external knowledge is not much different from the direct data input in general computers. The frequently used interview, spoken language record analysis, and machining copying knowledge acquisition methods all use this approach.

Knowledge Acquisition Methods by Using Editing Tools

The knowledge obtained must first be converted into an available form, before it can be loaded into the knowledge base. Knowledge acquisition thus needs two kinds of personnel involved, one being the experts in a specified domain, the other being the workers who have mastered knowledge engineering. In order to brief the work and reduce the workload involved, the system itself has to have the function of translation and compilation. Current knowledge acquisition tools are developed to achieve the above purpose, and the majority of knowledge acquisition tools use dialogue as their knowledge acquisition methods. For example, in the rule-based expert system for the diagnosis of infectious blood diseases, MYCIN, the relative experts act as teachers in knowledge acquisition, ask questions to the system MYCIN which is in student-mode, checks the answers that MYCIN gives out, find reasons for any error, and then amend the rules in the knowledge base. The knowledge acquisition system in PROSPECTOR, an expert system for mineral exploration developed by the Stanford Research Institute (Duda et al., 1979), also uses the knowledge editing tools. A knowledge acquisition system in teaching style requires a strong knowledge compiler, but general knowledge acquisition tools can only be described as ‘knowledge editor’. Existing system is relatively simple, and only has a part of the conversion functions. A powerful compiler should have the following functions besides its compilation function:

a. Knowledge management, including various editing features, such as knowledge retrieval, modification, deletion, insertion and so on.
b. Compatibility check-up, in order to ensure the consistency of the entire knowledge base; whether the new knowledge contradicts with existing knowledge in the knowledge base should be checked.
c. Integrity check-up, to ensure that the new knowledge meets the integrity constraint.
d. An explanation function, in the process of knowledge acquisition, to answer the users’ questions as much as possible, assist the experts to refine and debug the knowledge base, and give out the diagnostic report of the knowledge base.

Knowledge Acquisition Methods with Reasoning Capacity

This method has the ability to acquire new knowledge from the existing knowledge. The common characteristic is being able to educe full statements from special statements or knowledge, or infer unknown events from the observed events. These two types of knowledge acquisition methods (knowledge acquisition methods with no reasoning ability and by using editing tools) only need to convert the form of the information, without engaging in discovery. However, such knowledge acquisition methods need to collect various facts, and use them as general knowledge. The current knowledge acquisition methods with reasoning capacity include the following categories.

Interpreting Knowledge Acquisition
This is a method making use of logical inference. It introduces new knowledge from the existing

knowledge in accordance with reasoning rules. This interpreting knowledge acquisition system always serves by providing various explanations to the users. Once there has been initial knowledge and reasoning rules, users can continually ask the system questions, such as 'what kind of knowledge has been introduced?', 'why can a certain knowledge be added but others cannot?' and so on. The system will answer the users' questions by deduction and inference. Afterwards, users can decide whether or not to adopt the 'new knowledge' according to the system's explanation. In this way, new knowledge has not really been generated. Putting a theorem that can be proved will not expand the knowledge space, although it will reduce reasoning steps when proving theorems. This method can therefore make the inference chain shorter than the original steps, thereby enhancing the efficiency of the reasoning.

Inductive Knowledge Acquisition

Inductive knowledge acquisition is a large class of knowledge acquisition methods which obtain new knowledge by adopting inductive reasoning. This method induces conclusive and decision-making knowledge from the phenomena or data. It is more difficult than deduction for it is a synthetical process. According to the different inductive methods used, it can be divided into enumerating knowledge acquisition, analogy knowledge acquisition, associating knowledge acquisition and so on. Inductive knowledge acquisition attempts to solve the basic difficulties of automatic knowledge acquisition. No practical tools for inductive knowledge acquisition have appeared by now, however, once the inductive program is realized, it will bring fundamental changes to knowledge acquisition methods and level.

Knowledge acquisition is the most difficult but most important work during the development of KBS. It needs to be focused on and solved properly.

KBS EVALUATION

Why KBS Should be Evaluated

In the development of KBS, especially expert systems, much time is spent running test cases, observing the test results and collecting statistical information. There are at least three reasons for evaluating the KBS, which are:

1. Evaluation is the key to improving system design and its performance.
2. Evaluation will make users believe the availability of the system and thus make users accept the system.
3. Evaluation contributes to the development of Artificial Intelligence.

The development of KBS is actually a process with long-term feedback. It continuously put advice from users, domain experts and the constructors of the system into the construction of the system. At the same time, during evaluation, complex problems in the system will be exposed, such as the need to modify or expand the program, add, amend or delete rules in the knowledge base, and so on. Therefore, evaluation during the construction of the system is the key to improving the system design and its performance.

Whether a KBS is effective or not is symbolized by its scope in which the system is used as a tool, or whether it is accepted by the domain practitioners. Usually, some domain experts or practitioners will suspect the novel capacity of the system initially; thus, the system is challenged, and convincing evidence must be provided before users believe the diagnosis made by the system. In this sense, the acceptability of the users, namely, whether the users will accept the system or not, is the ultimate test.

Testing the system organizationally can provide objective criteria for comparing various methods and results. Some new methods or viewpoints may be introduced to explain the phenomena or

behavior observed in the experiments. Computer scientists consider the KBS as the source of observable phenomena. The evaluation provides data for the theory and practice analysis on Artificial Intelligence.

In the development of the system, the following issues must be continually considered.

1. Is the knowledge is appropriate; does it need to be amended or expanded?
2. Is the reasoning correct; are the results obtained correct?
3. Is the knowledge in the system consistent with the expertise?
4. Is it easy for users to use?
5. What kind of support equipment the users should provide?

In short, the system is always evaluated intentionally or unintentionally during its design and development.

Motivation

When conducting systematic evaluation, different people have different motivations. Who is undertaking the evaluation should first be confirmed, after which the motivation can be determined.

For system developers, evaluation might change their work. In order to ensure the propriety of the knowledge representation, the refinement of the knowledge base, the scope within which the system should be used, and any parts that might need improvement, system developers need to take into account the viewpoints and advice from users or experts, and confirm the availability of the system for users.

Domain experts normally evaluate a system statically and dynamically. In static evaluation, they compare the knowledge in the system knowledge base with their own expertise and look for integrity and consistency. In dynamic evaluation, by running test cases, they compare the conclusions inferred by the system with their own conclusions and check the knowledge base and inference strategies of the system.

For system users, users' views are very important to the success of the system. The ultimate standard for the success of the system is whether the system can be a really practical tool for users, rather than a tool for developers. Users' evaluation should therefore be included in the whole process of evaluation. By neglecting the ideas and requests of users, system developers will not be able to provide critical capacity and the system may only be partly available to users. Through users' evaluation, the system's performance and results can be identified. Also, necessary parts can be added and unnecessary elements can be deleted according to users' requests. At the same time, during the evaluation, users can get familiar with the system and learn to use it; they may also believe that the system has been improved and can now be put into use.

For AI researchers, the construction and evaluation of the KBS are mainly the exploration elements, not just an application. By comparing and evaluating the alternative systems and techniques, new ideas on development and new directions for exploration can be introduced. The development of AI will be promoted by abstracting new ideas and techniques. For example, in a slow system, researchers would want to find ways to improve the speed; a system with a large knowledge base would need effective storage management; a system with a poor user interface will require the development of user-oriented language and promotion of its ability for showing and explaining.

For psychologists, evaluation aims at exploring the human reasoning model. KBS attempt to encode the methods that human use (the methods for knowledge representation, logical inference and learning), and can be compared with human experts.

How to Evaluate

When a system is almost completed, every aspect of the system should be evaluated. The aspects that should be evaluated can be summed up as follows:

1. The quality of the recommendations and conclusions given out by the system.
2. The accuracy of the reasoning methods and control strategy.
3. The quality of the user interface.
4. The efficiency of the system.
5. The 'cost- efficiency' of the system.

It should be noted that the evaluation process is continuous. It starts from the design stage of the system and goes through the early stages of development in an informal way. With the progress of the development, it gradually turns into a formal process until it is completed. When the evaluation stops largely depends on whether users accept the system or not, or whether users' requirements for the accuracy and ability of the system are rigid or flexible.

The development process of the system and its corresponding evaluation are described below.

1. Top design, definition of the large scope goal.
2. Completion of prototype Mark-I, prove it can be realized.
3. Refinement of the system, and the informal evaluation.
    a. Run the informal test cases, the refined prototype Mark-II is formed through experts' feedback.
    b. Publicize Mark-II to customers and ask for comments.
    c. Revise the system according to the customers' feedback.
    d. Publicize the revised Mark-II, and go to (3) (b).
4. The structure evaluation for its performance.
5. The structure evaluation for users' acceptability.
6. Run a very long time in the prototype environment.
7. Further research to prove the system's availability in large practical scope.
8. Modify the program and allow the system's extensive configuration.
9. Publicize and sell with very strict updating plan.

The points above are illustrated as follows.

For the system designer, knowing clearly the motivation for developing the system is very important, and the large-scale objective must also be summarized.

Step 1 (initial design): system designers should know clearly what the measurement for the success of the system is and how to evaluate the system's failures and successes, which will affect the initial design and will help to develop a successful system. For example, if the ability to explain is crucial to users, the knowledge representation methods will be very important. The evaluation process therefore starts from the concept of the system.

Step 2: to prove the selected performance tasks are achievable, rather than attempting to prove that performance has achieved the experts' level. The system builders should make sure the knowledge representation is appropriate for the particular field, and that the prototype completes the sub-tasks in the specified domain very well. The evaluation during this period is informal, so a few simple examples are sufficient.

Step 3: the system runs informal test cases, and the system performance is checked, finally seeking the opinions of expert collaborators and potential users to confirm the major problems in the system development and to guide the work in the next period. This repeated process may last

for several months or several years. When to stop the process is dependent on the complexity of the problems, the flexibility of the knowledge and the availability of the special control strategies that are suitable for domain problems. Informal evaluation is one part of the repeated process.

Step 4: once the system can complete most test cases very well, more structure evaluation should be done, as the informal evaluation does not consider the actual availability of the system in the users' environment. If the formal random tests indicate the system can handle almost all the domain problems, then the evaluation in step 4 should be conducted. However, few systems other than MYCIN or PROSPECTOR have achieved step 4 at present. If the test evaluation which is randomly selected does not reach the level of experts, new exploration and knowledge requirements will be identified, and it will then be necessary to return to step 3 for further refinement.

Step 5: evaluation in the users' environment. The main problem of this evaluation is whether users can accept the system, or whether the system can meet the requirements of users. At present, only very few systems have reached this step. This evaluation also stresses the system's interactive ability with users and the hardware support it requires.

If the formal evaluation is satisfactory and users can accept the system, then continue to Step 6: run the system in the prototype environment for a very long time. Evaluation during this period achieves extensive testing experience through carefully observing the program's performance.

If the operation in the prototype environment is 'smooth', then large-scale availability of the system should be further studied (Step 7). Before the system is introduced to user groups, relevant parameters, such as the system's efficiency, cost-efficiency and so on, should be measured to determine the impact of the new environment on the system.

If new problems are found in Step 7, they should be solved before sale, including the modification and expansion of the program design, to make sure the system can run in different environments. All these become part of Step 8.

The final step is to release the system as a commodity. The main task in this step is to maintain the knowledge base and make strict plans to update the knowledge base.

APPLICATIONS OF KBS IN INTELLIGENT MANUFACTURING

In order to increase productivity in manufacturing, it is essential to develop and adopt new and innovative tools to help management and decision making within very complex environments. KBS are finding an increasing number of applications in intelligent manufacturing. They can be applied to almost all the manufacturing area, including design, process planning and scheduling, production control, and diagnosis. The following section describes the application of KBS in each stage of manufacturing.

Design Stage

It is widely accepted that the cost of a manufactured product is largely determined at the design stage. It is important to consider manufacturability in design stage, but tasks in the design stage have many different aspects which depend on the products to be designed, the manufacturing environments of the products and the creativity needed in the design, in short, the process of design is abstract and requires creativity. Consequently, KBS or expert systems have the smallest impact in this area. However, it is possible that an expert system that can emulate a designer's expertise in initiating a design for a product could be developed. To date, in the field of machine product design, most expert systems developed for design have not been used for generative type of design. They are often limited to variant type design or modification type design, in other words, they are used to assist the

designer in refining a design, or suggesting alterations. During this processes, when designers get a product to be designed, they first refer to the database, which contains the information about the products previously designed. They then select some candidate products that are similar to the product to be designed for feasible design solutions. In order to satisfy the design requirements, the design of the candidate products are modified and sometimes combined with other designs. Expert systems in this process are responsible to provide the designers with information and suggestion to perfect the design.

Process Planning

This stage follows the design stage and the development of KBS for process planning is considered much easier than that for design. The application of KBS in this area is normally divided into a generative planning method and a variant planning method. The principal functions of KBS for process planning are as follows:

- Selection of raw parts in stock
- Selection of the process and sequence of matching operations
- Selection of suitable machine tools
- Auxiliary functions (measuring instruments, fixtures, specifications)
- Selection of manufacturing parameter (such as setup, lead and process times; feeds and speeds).
- Generation of the text
- Output of the process planning

According to the manufacturing operation required, the generative method often involves a part surface to be manufactured. As variants for process planning such as machine tools, cutting tools, and fixtures are limited, KBS can be applied to perform tasks that are labor intensive and time consuming to optimize the modeling of all possible process solutions.

Generally, universal tools and production methods may be used for products with standard features and variant process can be used to determine the best solution for machining parts that have unique surface features. In this area, KBS play a very important role in improving flexibility and accelerating the search for an optimal process plan for manufacturing a product.

Manufacturing Adaptive Schedule

In dynamic manufacturing environments, the state of the manufacturing system is always changing. Many scholars therefore suggest the use of a scheduling strategy to replace the single scheduling rule, as such a scheduling strategy enables the scheduling rule of every decision-making time meet the requirements of the manufacturing systems (Tate, Drabble & Kirby). Scheduling with such a scheduling strategy is called adaptive scheduling. It is an efficient method of production scheduling and can choose appropriate scheduling rules according to the current working state and objectives of the manufacturing system. Moreover, it has dynamic scheduling capabilities and is very suitable for manufacturing systems with complex production systems.

In adaptive scheduling, the scheduling knowledge (or decision-making process) can be represented as production rules shown as follows:

Under objective, If condition, Then rule

Objective represents the current scheduling target of the manufacturing system to be executed; condition denotes the condition that the production of the manufacturing system should meet; rule is the scheduling rule. Scheduling knowledge realizes the mapping between specific scheduling problems and the scheduling rule.

In an adaptive scheduling system, KBS connect the scheduling experts or automatic knowledge acquisition technology with the scheduler. The knowledge of scheduling experts or the knowledge acquired by automatic knowledge acquisition technology is abstracted as the knowledge base.

For specific scheduling problems, KBS search for the appropriate scheduling rule, and then send it back to the scheduler.

In order to meet increasingly complex production requirements, much attention is paid to develop KBS for managing scheduling knowledge quickly and efficiently (Metaxiotios, Askounis, & Psarras, 2002). At present, there is only a prototype method for developing KBS. This method relies heavily on the ability of experts and their understanding of the knowledge, so it can not guarantee the stable quality and good performance of KBS.

Unified Modeling Language (UML) is a standard modeling language widely used in software engineering to describe software systems, visual processing and establish documents of software system products. Hakansson believes that UML can represent knowledge in a specific field, so it is feasible to use UML in knowledge engineering (Hakansson, 2001). Chung et al. developed a knowledge prototype system supporting the daily operation of small financial enterprises and achieved great success (Chung, 2006).

Diagnosis

KBS or expert systems are widely used in manufacturing process planning. The main functions of KBS are supervising complex production equipments and detecting the abnormal indications of the equipments; then locating problems as soon as the problems arise and estimating the causes of problems; finally determining suitable plans for the estimated causes. Therefore they can reduce production downtime by early location and resolution of problems. The reasons why KBS can be applied for diagnosis successfully are as follows:

- The knowledge base or the search space for a particular machine diagnosis is limited.
- KBS is able to explain the inference process and deal with fuzzy or ambiguous relationships.
- The inference process of KBS is adaptable. It can be applied for diagnosis of different machines just by modifying the rules in knowledge base.

The results of the diagnosis are often used to improve the maintenance and operation of the equipment. Therefore, for the equipment, KBS play a very important role in diagnosis, maintenance and operation.

Although KBS have been used successfully in this area, there remain many technical problems to be solved, such as the real-time data processing required in diagnosis, the generation of alternative plans, consideration of time-dependent factors and cooperation with other diagnostic systems.

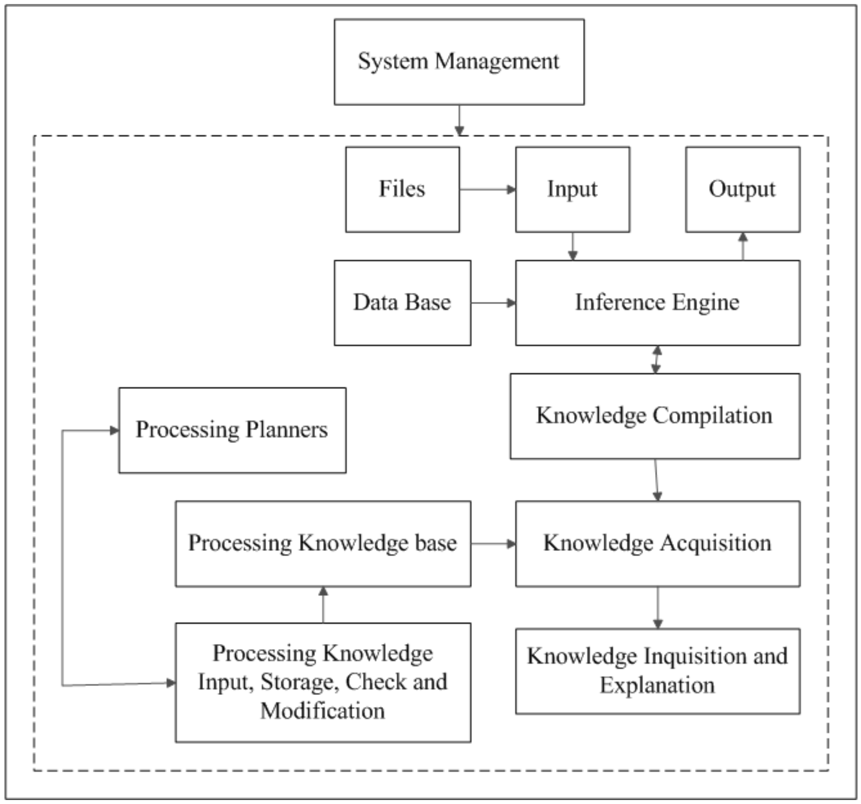
CASE STUDY

In this section, we will present a case study about the application of KBS in intelligent manufacturing—a computer-aided process planning expert system (CAPPES) for shaft design.

Background and Motivation

Computer-aided process planning (CAPP) has been a very popular research topic in the last three decades. Based on computer, it aims to mimic process planers to complete the design of process planning and greatly enhance the efficiency of process design. However, since the knowledge and decision-making methods of process design, which are very dependent on the practical experience of process planners, have no fixed pattern and cannot be described in unified mathematical models, it is very difficult to describe the process using traditional computer programs.

Figure 4. Working principle of CAPPES



The development of AI opens a new window for the development of CAPP. Since the 1980's, AI-based ES for CAPP has become one of the main research topics in manufacturing. On one hand, CAPPES is more flexible and able to deal with uncertain information and ambiguous expertise, which overcomes the limitation of traditional CAPP systems; on the other hand, CAPPES has the ability to dialogize and learn, which enable computers to really simulate process planners to design the process. CAPPES is different from a traditional CAPP system both in its working principle and basic structure. A traditional CAPP system normally has only two key components: the parts information input module and the process planning module. The process planning module, the core of CAPP system, includes knowledge and decision-making methods for process design. When product information is provided, the system will call the corresponding programs and generate the process plan. When manufacturing environment changes, the system programs need to be amended. This is difficult for users to manage and hence the system thus has limited or poor adaptability.

CAPPES is composed of information input module, knowledge base and inference engine, its working principle is shown in Figure 4. The knowledge base and inference engine is independent of each other. Unlike traditional CAPP system generating the process plan while running corresponding programs, CAPPES would visit the knowledge base frequently based on the input product information, search rules that can deal with the current states of parts through knowledge base according to the control strategies of inference engine, and then execute these rules and record conclusion of each executed rule in

sequence until the processing of parts reaches an end. Ultimately, the record includes the required machining process planning for the parts. Based on knowledge structure, CAPPES is organized into a three-tier structure of data, knowledge and control. The knowledge base is separated from each other, which increases the flexibility of CAPPES. Once the manufacturing environment changes, CAPPES can be updated by modifying the knowledge base and adding new rules in order to adapt to the new environment.

CAPP is the bridge connecting CAD and CAM and is also the central link in the realization of CIMS. AI based ES provides favorable technologies for the integration of CAD and CAM. CAPPES can obtain parts information directly from the database of CAD systems through AI interfaces, which in favor of the combination with CAD systems. Some even have considered the intelligent interfaces for production management systems, which creates favorable conditions for the realization of CIMS.

The next section will present the architecture of a CAPPES for shaft and its development, including the knowledge base, data base and inference engine.

Architecture of CAPPES for Shaft Design

The structure of CAPPES for shaft design is shown in Figure 5, the CAPPES was designed using modular software design method. The following sections will discuss the basic modules of the CAPPES.

Development of CAPPES for Shaft Design

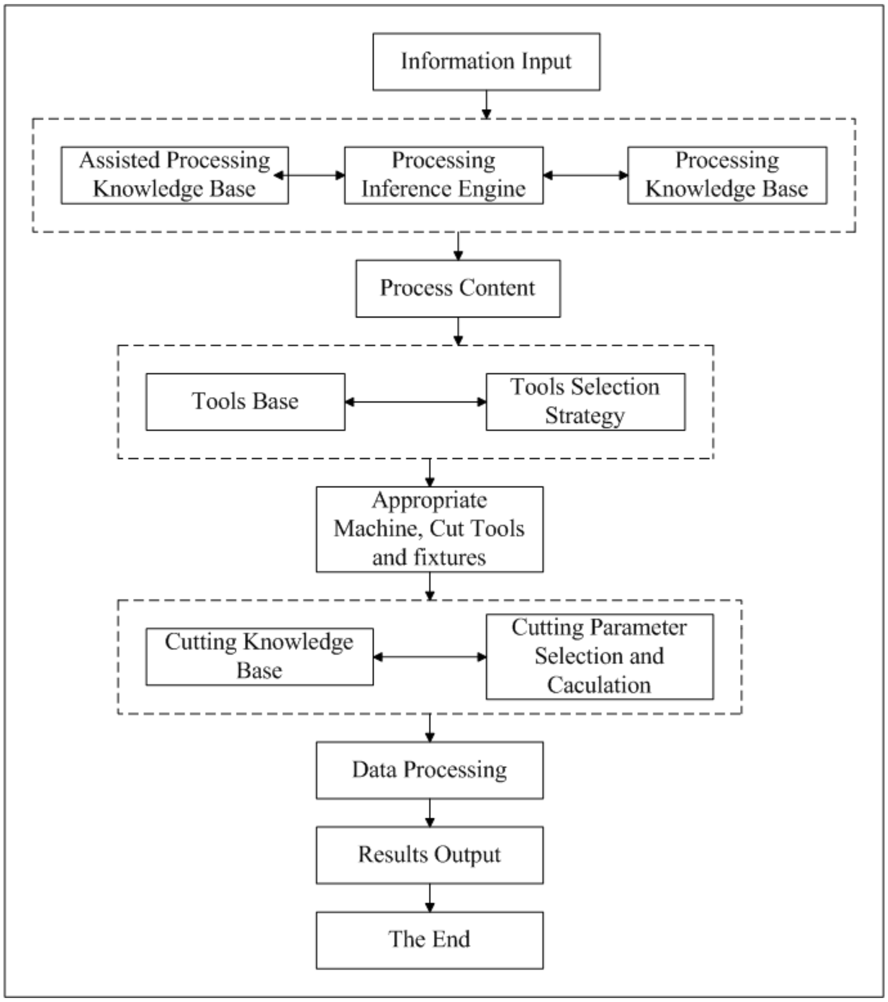
CAPPES for shaft parts should take into account the following factors, including the technical requirements of outer and inner surface, the selection of machine tools, fixtures, and cutting tools, and the decision of cutting parameters. In order to work out practical machining process, CAPPES needs to have comprehensive knowledge base and database management system. Moreover, it needs to be able to generate and formulate corresponding rules and obtain corresponding machining parameters in accordance with the requirements of the process structure.

Process Knowledge Base

There are 22 rules in the process knowledge base of CAPPES, some of which are shown as below. Each one can be considered as a table with three elements, including the rule name, the IF part and the THEN part.

```
Rule 1 IF   object is shaft
            parts
       AND  no center hole
       THEN process 1: turning
            hold the workpiece
            turning the right
            terminal face
            turn and cut off
            turning the right
            terminal face
Rule 2 IF   object is shaft
            parts
       AND  with center hole
       THEN process 1: turning
            hold the workpiece
            turning the right
            terminal face
            drill centers
            turn and cut off
            turning the right
            terminal face
            drill centers
Rule 3 IF   object is shaft
            parts
       AND  accuracy IT12-IT11
            and roughness Ra
            50-Ra 12.5
       AND  no function slot
       AND  process 1: turning
```

Figure 5. Overall structure of CAPPES for shaft design




```
THEN process 2: turning
     hold the workpiece
     rough turn the left
     outer circle
     chamfering
     turn around and
     then hold the work-
     piece
     rough turn the
     right outer circle
     chamfering
.
.
.
```

```
Rule 22      IF   object is shaft
                  parts
            AND   accuracy IT 6-IT 5
                  and roughness Ra
                  0.1-Ra 0.006
            AND   with keyway
            AND   process 1: turning
            AND   process 2: turning
            AND   process 3: mill
            AND   process 4: heat
                  treatment
            AND   process 5: grind
                  centers
            AND   process 6: grind
            THEN  process 7: abrading
```

The rules form the process knowledge base for CAPPES are stored in a table named RULES (rule 1^rule 2^……^rule 22). Each rule is an element or sub-table of RULES.

Database

Database of CAPPES is compromised of the current state of CAPPES, facts informed by users and facts obtained by inference. It is formed as a table named FACTS.

For example, if we know a fact that the object is a plain shaft, accuracy IT 10-IT 9 and roughness Ra 6.3-Ra 1.6, accuracy IT 7 and roughness Ra 0.8-Ra 0.4, center-hole, function-slots at both ends, and with keyway. The fact can be transferred in LISP language and stored in computer data base as below:

```
(SETQ FACTS
   (The object is shaft parts
   Plain shaft
   Accuracy IT 10-IT 9 and
   roughness Ra 6.3-Ra 1.6
   Accuracy IT 7 and roughness
   Ra 0.8-Ra 0.4
   Center hole
   Function slots at both ends
   With keyway))
```

Inference Engine

Inference engine is a very important part of CAPPES and CAPPES adopts forward reasoning strategies. Users first store a batch of facts into the database, inference engines work with these facts as the following steps.

1. Inference engine matches the facts provided by users with the premises or the IF parts of rules in knowledge base.
2. Add the conclusions or the THEN parts of rules which have been executed for their premises or IF parts match the facts provided by users to the database, therefore, the facts in the database increases.
3. Repeat the above two steps a) and b) with all the facts in the updated data base until the final conclusions are obtained or no more new facts are added into the data base.

Macro language LISP is used in CAPPES for shaft design to implement the above functions; some of the programs are described below.

1. Function for extracting the premise-facts of rules.

```
(CDADR RULE)
```

2. Function RECALL is defined to see whether the premise-facts of rules are in the database FACTS yet. If FACT is in FACTS, answer FACT, otherwise answer NIL. RECALL is defined as below:

```
//Define function RECALL and
variable FACT
(DEFINE (RECALL FACT)
//If FACT is an element of
FACTS, answer FACT or answer
NIL
(COND ((MEMBER FACT FACTS)
FACT) (T NIL)))
```

Forward inference aims to expand the database FACTS as much as possible by using the knowledge in knowledge base. The function RECALL defined above can only determine whether a single fact is in the database FACTS or not. As a rule may have several premise-facts, a new function TESTIF, therefore, is needed to determine if all the premise-facts of a rule are in the database FACTS or not. It needs to be able to testify automatically if all the premise-facts of a rule are in the database. Of all the premise-facts of a particular rule are in the data base, answer True otherwise answer NIL. TESTIF is defined as below.

```
//Define function TESTIF and
variable RULE
(DEFINE (TESTIF RULE)
//Process function and pro-
cess variable IFS
(PROG (IFS)
//Set IFS as a table for
premise-facts of RULE
(SETQ IFS (CDADR RULE))
LOOP   //Recycling logo
//IF all the premise-facts
are in the FACTS, return
True
(COND ((NULL IFS) (RETURN
True))
//Is the fact before in the
FACTS?
((RECALL (CAR IFS)))
//If one of the premise-
facts is not in FACTS, re-
turn NIL
(T (RETURN NIL)))
//Set IFS as its left tail
(SETQ IFS (CDR IFS))
//The next circle to see
whether the next fact is in
the FACTS
(GO LOOP)))
```

3. A new Function REMEMBER is defined to add the newly confirmed facts NEW which are from users or the reasoning process to the data base FACTS. If the fact NEW has been in the database, return NIL; otherwise, add the fact NEW into the database and return NEW. REMEMBER is defined as below.

```
//Define function REMEMBER
and variable NEW
(DEFINE (REMEMBER NEW)
// If the fact NEW has been
in the data base FACTS yet,
return NIL
(COND ((MEMBER NEW FACTS)
NIL)
// add the fact NEW into the
data base FACTS and return
NEW
(T (SETQ FACTS (CONS NEW
FACTS)) NEW)))
```

The function REMEMBER can add only one fact into the database FACTS once, and the THEN part of a rule may have a number of facts. Therefore, when a rule is executed, a new function is needed to put all its conclusion-facts into the data base one by one. We name the function USETHEN, whenever a new fact is added into the data base, print out 'RULE (the name of the rule) DEDUCES FACT (the name of the fact)'. If at least one fact is added, return T. USETHEN is defined as below.

```
//Define function USETHEN and
variable RULE
(DEFINE (USETHEN RULE)
//Process function PROG
and process variable THENS
SUCCESS
(PROG (THENS SUCCESS)
//Set THENS as a table for
the conclusion-facts of RULE
(SETQ THENS (CDADDR RULE))
LOOP        //recycling logo
//Set all the facts as the
parameters of function
REMEMBER
(COND ((NULL THENS) (RETURN
SUCCESS)) (RETURN SUCCESS))
// Set the first fact as
the parameter of function
REMEMBER
((REMEMBER (CAR THENS))
//If the fact is added into
the database successfully,
print out 'RULE (the name of
the rule) DEDUCES (the name
of the fact)' and return T.
(PRINT 'RULE (CAR RULE)
DEDUCES (CAR THENS)')
(SETQ SUCCESS T)))
```

```
//Set THENS as its left tail
(SETQ THENS (CDR THENS))
(GO LOOP))) //The next cir-
cle to judge the next fact
```

The determination of whether the whole premise-facts of a rule are in the database and putting the conclusion-facts of the rule into the database are executed separately. Here, a new function TRYRULE is defined to put the conclusion-facts of a rule into the data base automatically when all the premise-facts of the rule have been in the data base. Moreover, when all the conclusion-facts have been in the FACTS while not all the premise-facts are in the FACTS, return NIL; or when all the premise-facts have been in the FACTS while not all the conclusion-facts are in the FACTS, return T and print out 'RULE (the name of the rule) DEDUCE (the name of the fact)'. TRYRULE is defined as below.

```
(DEFINE (TRYRULE RULE)
(AND (TESTIF RULE) (USETHEN
RULE)))
```

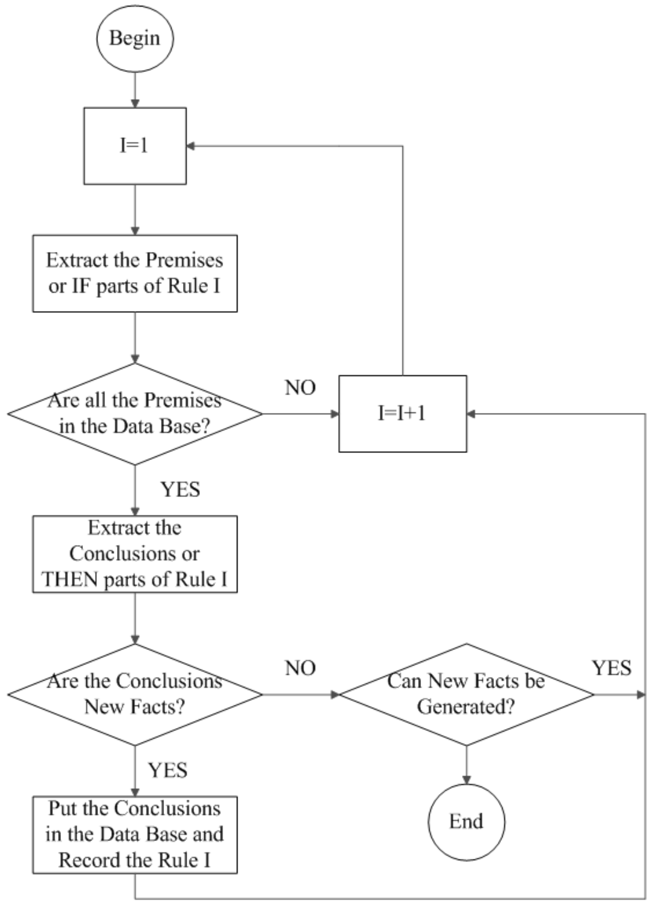
Function TRYRULE is trying to test a rule based on the current facts in data base FACTS. However, rules are more than one, in order to automatically scan the table of rules, function STEPFORWARD is defined. When the trial for a rule is unsuccessful, it will automatically turn to another rule until a suitable rule is found. Then add the conclusion-facts of the suitable rule to the data base FACTS and print out RULE (the name of the rule) DEDUCE (the name of the fact). If no suitable rules are found, return NIL. STEPFORWARD is defined as below.

```
//Define the function STEPFOR-
WARD
(DEFINE (STEPFORWARD)
//Process function PROG and pro-
cess variable RULELIST
(PROG (RULELIST)
//Set RULELIST as RULES
(SETQ RULELIST RULES)
LOOP //Recycling logo
//Scan all the rules, if no rule
can be used, return NIL
(COND ((NULL RESULT) (RETURN
NIL))
//If the first rule is executed
successfully, return T; or set
RULELIST as its left tail
((TRYRULE (CAR RULELIST)) (RE-
TURN T)))
(SETQ RULELIST (CDR RULELIST))
   (GO LOOP)))   //Turn to the
recycling
```

When the process function PROG is executed, NULL is assigned to its variable RULELIST, and the SETQ clause sets the RULELIST as the table of RULES, enters the recycling logo LOOP, finally turn to the condition function clause COND. The first clause of COND will not be executed here because RULELIST is not an empty set at first. The second clause of COND decides whether the trial for the first rule is successful or not, if it succeeds, return T and exit; or turn to next SETQ to set RULELIST as RULES' left tail and then go to the loop. If the RULELIST has been empty at this time, return NIL and quit, otherwise, repeat the above processes.

Function STEPFORWARD tries the RULES in knowledge base based on the facts in database FACTS and expands the FACTS. However, the forward reasoning strategies are able to continually try the RULES based on updated facts in database FACTS and expand the database over and over again, which means expand FACTS according to updated FACTS until FACTS can not be expanded. Therefore, a new function DEDUCE is defined to call the function STEPFORWARD automatically and repeatedly. If function STEPFORWARD is not executed successfully once, return NIL; otherwise, return T. DEDUCE is defined as below.

Figure 6. The process of forward inference




```
(DEFINE (DEDUCE) //Define func-
tion DEDUCE
//Process function PROG and pro-
cess variable PROGRESS
(PROG (PROGRESS)
LOOP //Recycling logo
//IF STEPFORWARD is executed
successfully, return T
(COND ((STEPFORWARD) (SETQ PROG-
RESS T)
//Otherwise, return PROGRESS
(T (RETURN PROGRESS)))
(GO LOOP) //Turn to LOOP
```

When the process function PROG is executed, NIL is assigned to its variable PROGRESS. The first clause of condition function COND is to check whether STEPFORWARDD can be executed successfully or not. If STEPFORWARDD is not executed successfully, run the second clause of the condition function COND and return PROGRESS. If STEPFORWARDD is executed successfully, assign T to PROGRESS and turn to LOOP, and data base FACTS has been expanded yet.

In conclusion, with the knowledge table RULES in knowledge base, the fact table FACTS in database and reasoning function DEDUCE which continually expands the FACTS by using RULES and current FACTS, CAPPES can use forward inference strategies to make suitable process planning for shaft design. The forward inference is shown in Figure 6.

SUMMARY

Knowledge-based systems are typically domain-specific and tend to operate by utilizing heuristic problem-solving approaches. This is a branch research area of Artificial Intelligence, which has achieved considerable success in the development of KBS. Though it is very difficult to establish a perfect intelligent KBS, KBS have made breakthroughs in many important applications. In conclusion, KBS have proved their value in a remarkable number of applications.

In this chapter, an overview of the KBS has been presented. The history and features of KBS were briefly introduced and compared with human intelligence, and a classification of the problems KBS can solve was given. The process of building a KBS and knowledge engineering was discussed, together with techniques in knowledge engineering, including technologies about knowledge representation, knowledge acquisition, and inference strategy and conflict resolution. The application of KBS in manufacturing intelligence is also discussed and a case study of KBS for process planning is presented. Finally, the problems related to systems' evaluation were reviewed, including the questions of why, what and how to evaluate.

In spite of the remarkable number of successful applications of expert systems in different areas, there are many remaining problems to be solved. Most of the KBS are difficult to verify and validate, have limited explanation capabilities and are not as robust and flexible as expected. Moreover, they do not work efficiently enough, and their problem solving capacity needs to be further improved.

One reason for this is that current expert systems mainly simulate experts in a specific field to solve problems and do not pay attention to collaborative problem solving among different experts. In the real world, collaborative problem solving is more popular and common; problem solving in a specific area is restricted to a very narrow domain. Therefore, although expert systems succeed in simulating human experts in a particular field, they still have limitations to solve difficult problems in the real world.

Secondly, current expert systems are becoming larger and more complicated, so large-scale expert systems should be divided into a number of small, relatively independent expert systems, and expert systems in different areas should be combined to solve problems collaboratively. However, the theory and practical technology on distributed AI are still at early research stage.

Only when a solution to the problem of distributive and collaborative problem is found can the 'vulnerability' of the system due to limited knowledge and problem solving methods of a single expert system be overcome. Once that happens, KBS will become more reliable and flexible.

REFERENCES

Barr, A., & Feigenbaum, E. A. (Eds.). (1981). *The Handbook of Artificial Intelligence* (*Vol. 1*). Los Altos, CA: Morgan Kaufmann.

Bobrow, D., & Winograd, T. (1977). An overview of KRL, a Knowledge Representation Language. *Cognitive Science*, *1*, 3–46.

Brachman, R. J., & Levesque, H. J. (2004). *Knowledge Representation and Reasoning*. Los Altos, CA: Morgan Kaufmann.

Chung, W. W. C. & Pak, J. J. F. (2006). A Case Study: Using UML to Develop a KBS for Supporting Business in a Small Financial Institute [J]. *International Journal of Computer Integrated*.

Cowan, J. D. (1990). Neural Networks: the Early Days. In Tourefzky, D. S. (Ed.), *Advances in Neural Information Processing Systems 2* (pp. 828–842). San Mateo, CA: Morgan Kaufman.

Cox, E. (1999). *The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using and Maintaining Fuzzy Systems* (2nd ed.). San Diego, CA: Academic Press.

Davis, R., & Lenat, D. B. (1982). *KBS in Artificial Intelligence*. New York: McGraw-Hill International Book Company.

Duda, R., Gaschnig, J., & Hart, P. (1979). Model Design in the PROSPETOR Consultant System for Mineral Exploration. In Michie, D. (Ed.), *Expert Systems in the Microelectronic Age* (pp. 153–167). Edinburgh, Scotland: Edinburgh University Press.

Durkin, J. (1994). *Expert Systems Design and Development*. Englewood Cliffs, NJ: Prentice Hall.

Feigenbaum, E. A., Buchanan, B. G., & Lederberg, J. (1971). On generality and problem solving: a case study using the DENDRAL program. In Meltzer, B., & Michie, D. (Eds.), *Machine Intelligence 6* (pp. 165–190). Edinburgh, UK: Edinburgh University Press.

Hakansson, A. (2001). UML as an approach to modeling knowledge in rule-based system [C]. In *Proceedings of ES 2001, the Twenty-first SGAI International Conference on Knowledge Based Systems and Applied Artificial Intelligent,* (pp. 187-200). Cambridge, UK: Springer.

Li, H. S., & Qing, L. F. (2003). *Battlefield Information System*. Beijing: Defense Industry Press. (In Chinese)

McCarthy, J. (1958). Programs with common sense. In [), London.]. *Proceedings of the Symposium on Mechanisation of Thought Process*, *1*, 77–84.

McCarthy, J., & Hayes, P.~J. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. *Machine Intelligence*, *4*, 463–502.

Medsker, L., & Leibowitz, J. (1994). *Design and Development of Expert Systems and Neural Computing*. New York: Macmillan.

Merdan, M., Terzic, I., Zoitl, A., & Favre-Bulle, B. (2005). Intelligent reconfiguration using knowledge based agent system. In *10th IEEE Conference on Emerging Technologies and Factory Automation, (ETFA)*, (Vol. 1, pp. 19-22).

Metaxiotios, K. S., Askounis, D. & Psarras, J. (2002). Expert Systems in Production Planning and Scheduling: A State-of-the-art Survey [J]. *Journal of Intelligent Manufacturing*, (S0956-5515), *13*(4), 253-260.

Michalewicz, Z., Schmidt, M., Michalewicz, M. & Chiriac, C. (2005). Case Study: an Intelligent Decision Support System. *Intelligent Systems, Vol. 20*, 8(4), 44-49.

Minsky, M. L. (1975). A frame work for representing knowledge. In Winston, P. (Ed.), *Psychology of Computer Vision* (pp. 211–277). New York: McGraw-Hill.

Newell, A., & Simon, H. A. (1961). GPS, a Program That Simulates Human Thought. In Billing, H., & Oldenbourg, R. (Eds.), *Lernemde Automatten* (pp. 109–124). Munich.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

Omlin, C. W., & Giles, C. L. (1996). Rule Revision with Recurrent Neural Networks. *IEEE Transactions on Knowledge and Data Engineering*, *8*(1), 183–188. doi:10.1109/69.485647

Rowe, T. (1988). Definition, Diagnosis and Origin of Mammalia. *Journal of Vertebrate Paleontology*, *8*, 241–264.

Shortliffe, E. H. (1976). *MYCIN: Computer-Based Medical Consultations*. New York: Elsevier Press.

Stefic, M. (1995). *Introduction to Knowledge Systems*. San Francisco: Morgan Kaufmann.

Stefik, M. (1982). The organization of expert systems: A tutorial. *Artificial Intelligence*, *18*, 135–173. doi:10.1016/0004-3702(82)90038-8

Tate, A., Drabble, B., & Kirby, R. (1994). *Intelligent Scheduling* (Zweben, M., & Fox, M. S., Eds.). San Francisco: Morgan Kaufman.

Turban, E., & Aronson, J. E. (2000). *Decision Support Systems and Intelligent Systems* (6th ed.). Englewood Cliffs, NJ: Prentice Hall.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*, 433–460. doi:10.1093/mind/LIX.236.433

Wang, W. D., Liu, X. H., & Liu, W. J. (2007). Development of KBS for Adaptive Scheduling Using Unified Modeling Language. *IEEE International Conference on Control and Automation, (ICCA), 7*, 914-918.

Wang, Z. M. (2006). *Knowledge Engineering and Expert Systems*. Beijing, China: Chemical Industry Publishing House. (in Chinese)

Winston, P. H. (1984). *Artificial Intelligence*. Reading, MA: Addison-Wesley.

Winston, P. H. (1992). *Artificial Intelligence* (3rd ed.). Reading, MA: Addison-Wesley.

Wu, H. Z. (1994). *Research and Practice of Mechanical Design Expert Systems*. Beijing, China: Railway Publishing House. (in Chinese)

Zadeh, L. (1965). Fuzzy Sets. *Information and Control*, *8*(3), 338–353. doi:10.1016/S0019-9958(65)90241-X

Zahedi, F. (1993). *Intelligent Systems for Business: Expert Systems with Neural Networks*. Belmont, CA: Wadsworth.

Chapter 3
Intelligent Agents and Multi-Agent Systems

Multi-Agent systems (MAS) are typical KBS and intelligent agents are viewed as extensions of KBS. Originating from the field of Distributed Artificial Intelligence (DAI), agent and Multi-Agent (MA) technology has been at the forefront of research in the last decade (Nilsson, 1998). Since the late 1980s, researchers have applied agent technology to perform tasks, and it is considered a promising paradigm for intelligent manufacturing (Shen & Norrie, 2001). In the 21st century especially, the manufacturing industry has become more and more competitive in a market that is frequently changing. Manufacturing systems should therefore move to support product innovation, global competitiveness and rapid market responsiveness. Recent new developments in agent and MA technology have brought new and interesting possibilities (Jennings & Wooldridge, 1998), researchers have been trying to develop and apply agent technology for supporting intelligent manufacturing, and there have been many projects in agent-based intelligent manufacturing. The basic theory and applications of agent and MAS are introduced in this chapter. The recent development of agent and MAS is reviewed, and the current research level of MAS is also summarized. Finally, the fundamentals of agent technology including communication and interaction, collaboration and behavior coordination, are presented.

DOI: 10.4018/978-1-60566-864-2.ch003

INTELLIGENT AGENTS

'Intelligent' can be utilized as both a noun and an adjective. If considered as a noun, it refers to all the mental activities human beings are capable of, such as feeling, cognition, memorization, relating, calculating, reasoning, judging, decision-making, and summarization. If used as an adjective, it carries the meaning of being human-like, smart, flexible, self-learning, self-organizing, self-adaptive, and autonomous.

The research with regard to the theory of intelligence is divided into two aspects. The first is to conduct direct research for the forming and working mechanisms of intelligence, which is classified as natural "intelligent" theory and has mainly been researched by physiologists and psychologists. The second aspect is to explore methods to stimu-

late and expand intelligence artificially, which is classified as artificial "intelligent" theory and has mainly been used by engineers. In the first aspect, 'intelligent' is often used as a noun because it mainly explores the mechanisms of mental activity. 'intelligent' is used as an adjective in the second aspect because what we really care about is whether functions of AI are better than functions of natural intelligence.

Definition and Properties

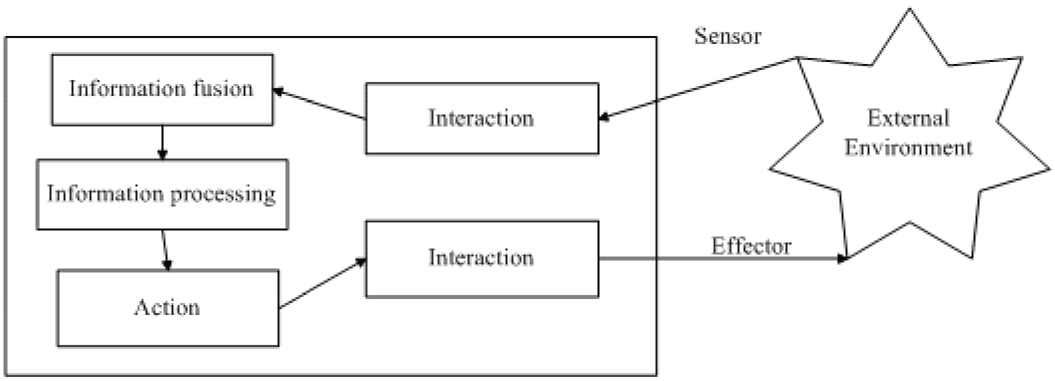
The development of Intelligent Agents are a natural result of the development of distributed AI technology and network technology. Intelligent agent and MA have been hot topics in the area of network-based distributed AI. References even claim that agent technology has been a significant breakthrough in network software.

The concept of agent can be traced back to the year of 1977 when Hewitt published the article 'Viewing Control Structures as Patterns of Passing Messages'. In this paper, an object was first defined as self-compatible, interactive and capable of parallel processing. The object has closed internal states and is able to conduct message-exchange and feedback with other similar objects (Hewitt, 1977). The term 'Agent' can first be found in the book 'Society of Mind', published by Minsky in 1986. In this book, Minsky introduced the concepts of 'Society' and 'Society Behavior'. Every individual is a part of the society, but different individuals can solve problems through coordination and competition despite the contradictions between them (Minsky, 1986). In 1994, Minsky elaborated his ideas on agent. From his perspective, agent is an entity that possesses unique skills. For computers, agent is referred to as a machine that can accomplish certain tasks and operators do not have to know its working mechanisms. This machine is treated as a functional black box. In this way, Minsky revealed two important attributes that are essential to an agent, namely, sociality and intelligence (Minsky, 1994). In 1996, Franklin and Graesser presented a paper at the third international workshop on agent theories, architectures and languages called 'Is it an Agent, or just a program?: a taxonomy for autonomous agents'. In this paper, the authors collected a variety of definitions on the intelligent agent, and drew out their own definition. They defined an autonomous agent as a system situated within a part of an environment that senses the environment and acts accordingly, over time, in pursuit of its own agenda, which affects what it senses in the future (Franklin & Graesser, 1996).

Intelligent agent is an abstract noun that can represent all intelligent entities whether they are of natural intelligence or artificial intelligence. It is therefore used to describe a wide range of entities, such as human beings, robots, intelligent devices, and intelligent software. In a certain environment, an intelligent agent can sense the environment through sensors and affect the environment through effectors. The working mechanisms of intelligent agents are shown in Figure 1. It is worth mentioning that an intelligent agent cannot exist alone in an environment, but rather works coordinately with many other agents through communications and message-exchanging. Nevertheless, every individual agent can function actively and autonomously, which makes distributed AI possible, and has enormous practical value in such areas as parallel programming, computer communications, network management and control.

Agent research is currently extremely popular in many areas, such as in AI and in computer science (Jennings, 2000). The theory of agent and MAS, the architecture of agent, communications and message-exchanging between agents in a MAS, and the language of agent have all been key areas of interest. Scientists even put forward a new definition for AI based on intelligent agent. It is defined as a branch of computer science and its goal is to create agents that are capable of undertaking certain intelligent behaviors. In a special report at IJCAI'95, Hayes-Roth from

Figure 1. Structure of agents



Stanford University defined an intelligent computer agent as both the initial and ultimate goal of AI (Hayes-Roth, 1995).

There is still no universal concept for the definition of the term 'agent', and almost all the current research for agent is still only at the initial stage; many important issues have yet to be resolved. All these factors have contributed to the fact that no universally acknowledged definition for agent yet exists. In a broad sense, a wide variety of entities are regarded as agents, including human beings, physical robots and software robots, but in a narrow sense, only software robots are called agent or software agent. It represents mobile computing entities that can perform a set of operations as active service. The term 'active service' has two meanings:

1. Active adaptation, which means that the agent should be capable of gathering knowledge about operating objects, users' expectations and preferences, and using this knowledge in future operations.
2. Active implementation, which means that the agent should be able to perform operations without the user instructions if the current state fits certain conditions. 'Mobile' is a term used to refer to the ability of an agent to access computing resources, communicate and coordinate with other agents. In this chapter, the agent discussed is limited to the definition of the software agent. The definition of agent given by Wooldridge in the paper 'Intelligent Agents, Theory and Practice' has been widely used by many researchers (Wooldridge & Jennings, 1995).

Definition 1: agent is used to describe a common hardware and software system, which has the following four characteristics (Maes, 1990).

1. **Autonomy.** Agent should be able to operate without interventions from users or other agents. It should also have control over its own behavior and internal states.
2. **Social ability.** Agent should be able to communicate with other agents (probably human beings).
3. **Reactivity.** Agent should be able to understand its surrounding environment and reactively respond to changes.
4. **Proactive.** Agent should not only be able to be responsive to the environment, but also should be able to display purpose-driven behaviors after receiving instructions.

Definition 2: Besides the characteristics discussed in Definition 1, agent should also possess unique features similar to human beings, such as knowledge, conviction, obligation, and purpose.

According to Shoran, agent is an entity comprised of various mental activities, such as conviction, capability, choice and commitment. Moreover, agent should have these features incorporated into its list of characteristics.

1. **Mobility.** Agent should be able to move freely on the Internet.
2. **Veracity.** False information should not be allowed to be transferred by agent.
3. **Benevolence.** If Agent doesn't have contrary tasks, it must do what it is asked to do.
4. **Rationality.** Agent should always work for a set goal. It should never stopping trying to achieve this goal.

The definitions above reveal the basic properties of agent. In fact, there are still many other properties for agent. However, it is extremely hard, if not impossible; to give complete list of definitions and it is not necessary to provide such a list. Agent should at least have the following attributes:

1. **Acting on Behalf of Others.** Agents should be able to act on behalf of others. They work for users, not for themselves. This is the most important attribute for agents. In addition, agents can pack all the other resources together and access these resources on behalf of users. In this case, agents serve as a bridge between resources and users.
2. **Autonomy.** Agent should be a computing entity that can work independently and has a certain degree of autonomy. This means that agent should be able to function properly with little or no outside intervention. It should also be able to independently carry out complicated operating procedures and solve problems in an environment that is unpredictable and dynamic. It should be able to allocate resources and services that suit the needs of users without the users' participation.
3. **Proactive.** Agent should be able to take proactive actions in accordance with previous commitments and exhibit purpose-driven behaviors. Agents on the internet, for example, should be able to roam over the whole network to collect and provide information users.
4. **Reactivity.** Agent should be able to interact with its surrounding environment and make proper response to related events.
5. **Social ability.** Agent should have a certain degree of social ability. 'Social ability' refers to the ability to conduct communications and message-exchanging with users, resources and other agents.
6. **Intelligence.** Agent should have a certain degree of intelligence, including a series of capabilities, such as pre-defined rules and self-learning AI reasoning machine. Agent should be able to comprehend users' resource requirements expressed in human language, for example; it should also be able to help users clear away language obstacles and get to know users' tastes and preferences; it should be able to predict users' intentions and realize these intentions for users.

Classification

Agent is a comprehensive term; it includes a series of different Agents. According to different criteria, Agents can be divided into several types.

According to Agents' functions, they can be classified into interface Agent (also known as user Agent or personal Agent) and software Agent (also known as task Agent)(Bradshaw, 1997).

According to Agents' characteristics, they can be divided into intelligent Agent, cooperative Agent, autonomous Agent and emotional Agent.

According to Agent's architecture, there are three types of Agents: cognitive Agent, reactive Agent and hybrid Agent. These three architectures will be described in the next section.

Architecture of Agents

There are mainly three architectures for Agent.

Cognitive Architecture

Cognitive architecture is based on the assumption of the physical symbol system brought forward by Newell and Simon. It includes a symbolic model and explicit expression for the world and environment, using logical or pseudo-logical reasoning for decision-making. A typical example of such a cognitive architecture is the STRIPS system developed by Fikes and Nilsson (1971). In 1987, Chapman theoretically proved that this architecture cannot be used to realize a real-time system and raised doubts about the effectiveness of symbolic AI (Chapman, 1987). This was why reactive architecture was then introduced.

Reactive Architecture

The characteristics of reactive architecture are that it involves nothing about the symbolic model to represent the world, and it does not have the symbolic reasoning mechanism in the architecture. A typical example for reactive architecture is the 'machine bug' developed by Brooks.

Hybrid Architecture

This type of architecture is a combination of the two architectures discussed above. The cognitive subsystem in this architecture that is responsible for overall planning and decision-making contains a symbolic model for the world. However, the reactive subsystem can make immediate response to emergencies without complicated reasoning processes. Under normal circumstances, the reactive subsystem has a higher priority than the cognitive system to ensure that the whole system can make an immediate response to important emergencies. The popular architecture for this type of Agent is layered architecture. Typical systems for this architecture include Touring Machines and InteRRap. The hybrid architecture for the Touring Machine system comprises a cognitive subsystem, a reactive subsystem and three controlling layers, which respectively is the responsive layer, planning layer and modeling layer. The two subsystems directly exchange messages with the interfaces to the outside environment. The three controlling layers can communicate with each other and exchange messages and a controlling frame is available to coordinate them in order to avoid potential conflicts. InteRRap is also a layered architecture, which includes a knowledge base and a set of controlling components. Controlling components are divided into four parts: the controlling component for the interfaces connecting the world; the behavior-based controlling component; the planning controlling component; the coordinating controlling component. The controlling component for the interfaces connecting the world and its corresponding world model knowledge base lie at the very bottom of this system. These two parts are responsible for communications and message-exchanging with the environment. The behavior-based controlling component controls Agent's basic responsive capabilities. The planning component includes a planner to answer requests from the behavior-based controlling component. A plan for an individual Agent will be created along the answering process. The coordinating component can create a coordinated plan that suits the needs of more than one Agent.

Reactive architecture agent system (also known as reactive Agent) only responds to changes in the environment and messages from other agents, but cannot to change its internal states. It only operates in accordance with triggering rules and pre-defined plans. This type of system can respond promptly to changes in the environment and outside messages, but its flexibility is unsatisfactory. The cognitive architecture Agent system can use symbols and logic to create, renew, evaluate, choose and implement suitable plans in order to realize

its purpose. It enjoys a high degree of flexibility and intelligence, but its swiftness of response to changes in the environment is a problem. According to current research results, the cognitive Agent system has held a predominant position in the research for Agent architecture, and the research and application of reactive systems are still at an initial stage. It is suggested that neural networks, genetic algorithms and machine-learning should be adopted to solve the problems that exist in this type of system. Some scientists think that a combination of these two types of architecture provides the best choice.

BASIC THEORIES OF MULTI-AGENT SYSTEMS

Introduction

A Multi-Agent System (MAS) is referred to as a set of agents that can conduct network computation. Generally speaking, every individual agent is regarded as a physical or abstract entity. In a network-based distributed environment, every agent is an independent entity that can exert impact on external environment and its own. It can also change the environment and respond to changes in the environment. More importantly, it can communicate with other agents, exchange messages and work in coordination to accomplish a task.

The definition of MAS is as follows: MAS is a distributed intelligent system that can solve problems and alter its own behavior in accordance with changes in the environment (Ferber, 1998). A MAS should also be able to communicate and cooperate with other agents to accomplish a task. This type of system should be able to simulate the working mechanisms of a large human community and use human beings' problem-solving methods to untangle complicated issues of common concern. The intelligence level of a MAS in open distributed environment is dependent on the AI in the system. The AI controls how a group of dispersed and loosely coupled intelligent entities work cooperatively in the distributed environment to solve problems. In a MAS system, individual agent has a high degree of autonomy and is capable of reasoning, planning and communicating in response to its own knowledge and surrounding events. Different agents are logically independent but can coordinate to find solutions to problems by sharing knowledge, tasks and intermediate results. 'Dispersed' means that data and knowledge are logically or geographically dispersed. 'Loosely coupled' means that individual agents will put more emphasis on computing or data-processing, rather than on communication.

Distributed Artificial Intelligence (DAI) is an important branch of AI which explores how a group of concurrent and autonomous agents work cooperatively and interactively to solve problems. Traditional AI, which puts more emphasis on using one agent to solve a single problem, is the basis for the development of DAI. The reasons for the development of DAI are as follows:

1. Human beings' intelligence not only manifests itself in an individual's intelligent behaviors, but, more importantly, in his performance in different organizations. DAI can therefore better describe human intelligence than traditional AI.
2. The resources and knowledge of a single intelligent system are very limited. By getting a group of Agents to work in coordination and interactively, the technology of DAI substantially enhances AI's problem-solving capabilities.
3. The development of AI and neural network, and the emergence of the blackboard system and the Actor framework, have all played a part in the development of DAI.
4. Compared with traditional AI, DAI has many advantages, such as high efficiency, good robustness, good extensibility and excellent adaptability.

5. The problems in the real world are extremely complicated, but the emergence of DAI has provided an excellent way to solve all these complex problems.
6. The rapid advancement of computer hardware technology and large-scale network technology has made the research of DAI necessary and has also given a major boost to the development of DAI.

The research of DAI began in the late 1970s, with Distributed Problem-Solving (DPS) and MAS as its two main directions. DPS mainly explored how to divide a certain problem into different parts and assign each part to an Agent to be processed. MAS, however, put more emphasis on coordinating a group of autonomous or semi-autonomous agents to get problems solved and focused on how different agents coordinated their respective knowledge, purposes and strategies.

The research of DPS is the origin of DAI and held a predominant position in the early days of DAI research. Since the research of DAI was based on the development of traditional AI, it was a natural choice for early scientists to explore DAI from the perspective of traditional AI, which created a distributed system to solve certain problems. When the existing technologies could not solve specific problems, however, scientists took advantage of the rapidly developing network technologies and combined them with traditional AI methods to unravel these problems. This combination produced the so-called DPS. Meanwhile, scientists also wanted to enhance the system's robustness and problem-solving capability with the development of DPS.

The focuses of MAS research are Agent itself and the coordination and cooperation between different Agents. The definition of Agent varies substantially in different systems, ranging from the fine-grained processing units to the coarse-grained knowledge base, expert systems, and robotics. In MAS, Agents both work independently and cooperatively.

The research into cooperation and coordination between groups of autonomous Agents is at the heart of MAS research and all other research directions are essentially based on that research. Agents should figure out how to coordinate their knowledge and purposes so that they can act together and find solutions to problems of common concern. An individual Agent in MAS may have only one goal and it can also have more than one goal.

Since the MAS can better display human beings' intelligence, has better flexibility and adaptability and is more suitable for open and dynamic environment, people have attached more and more importance to it. Due to the rapid development of the Internet, MAS research is no longer confined to traditional DAI. It is predicted that DAI will only be a topic in future MAS international conference. Previous MAS research is mainly cooperation-oriented, but the tendency now is to explore common problems from the perspective of a single Agent. The main task here is to figure out what architecture and capabilities a single Agent should have in order for it to function independently and communicate with other Agents in a time-limited and open environment.

Architectures of MAS

Generally speaking, MAS has two main architectures (Genesereth, 1993). The first is pure distributed architecture, in which all Agents in the system can share information and knowledge and every individual Agent is capable of conducting message-exchanging and communications. When an individual Agent needs the services of other Agents, it only needs to make a request to them. After receiving the incoming requests, other Agents should first assess the request and then exchange messages with the Agent. In this way, a contractual commitment will be established to accomplish tasks. This type of architecture is suitable for small systems, since when the system is very large-scale, the communication pressure

for the system will be enormous and the system structure will be very complicated. Though this architecture has very good regional autonomy, it is very hard to realize global optimization, which is very important for MAS. The other type of MAS architecture is the federated architecture. This architecture has a coordinating mechanism based on Mediator (Gaines & Norrie & Lapsley, 1995). The Mediator can gather a group of Agents into a package of Agents, and the individual Agent in every package only needs to communicate with its Mediator. The Mediator is mainly responsible for the coordination between Agents within the package and for message-exchanging between different packages. The type of MAS architecture reduces the complexity of communications and control in the first architecture and is capable of solving complicated problems. But the Mediator must shoulder the responsibilities of helping the system to figure out partners, delivering messages correctly and dealing with potential problems.

Foundation of Intelligent Physical Agents

The Foundation of Intelligent Physical Agents (FIPA) is a non-profit international organization (Http://www. fipa.org/specs). The main purpose of this foundation is to promote the development of technical standards and support Agent-based applications. The architecture of FIPA includes: a series of definitions and norms, architecture guidance for the creation of Agent platform, mapping specifications of abstract architectures and specific applications, specifications for the management of agents and Agent platform, and a series of non-binding appendices for the realization of FIPA norms.

FIPA architecture gives a general description of the components of the FIPA Agent system and the relationship between these components. From the application perspective, the planning of an abstract architecture relies heavily on the analysis of existing and recommended mechanisms (such as CL, agreement and network services) so that an effective and strong architecture is created. The architecture of FIPA has addressed the following questions:

1. How to create replacing mechanisms for a specific function, such as a mechanism allowing multiple transmission of a message.
2. How to combine existing or upcoming technologies with FIPA specifications. This includes all kinds of technologies, such as XML, SMTP, Java, ACTIVE OBJECT, CORBA, directory model, WEBSERVER, e-commerce and all kinds of message-transmitting methods.
3. How to define and describe interoperability between different layers, such as the relationship between different Agents, the relationship between Agent and Agent platform and so on.
4. How to define a consistency model and conformance testing so that the consistency of FIPA specifications is ensured.

FIPA specifications are only the first step towards Agent standardization. They have not described the internal architecture of Agent, nor have they elaborated on how to realize it, but they have specified interfaces essential to interactive behaviors. FIPA specifications mainly define the following aspects:

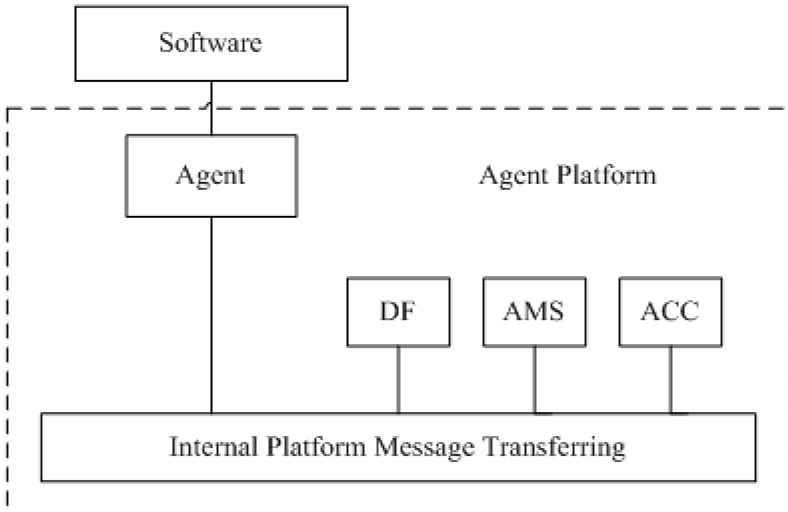
Agent Communication

Agent communications include interactive behaviors between agents, such as consultation, collaboration and message-exchanging.

Agent Management

Agent management mainly specifies the Agent platform. The FIPA recommended model (shown in Figure 2) provides a standardized framework for the existence and operation of Agents. In Figure

Figure 2. The FIPA reference model of agent platform



2, the directory facilitator, AMS and ACC are all special Agents responsible for managing other Agents. DF provides a 'Yellow Pages' service for other Agents. AMS and ACC are responsible for conducting internal communications of agents by dint of the FIPA Agent communication language. ACC is also responsible for providing interaction interfaces within platform or between platforms. In order to be compatible with FIPA, an Agent platform should at least support IIOP (Internet Inter-ORB Protocol). Internal platform messaging (IPM) provides relaying services for Agent. IIOP protocol and IPM together constitute the Agent platform, which is a mandatory and standardized component of this model.

The Integration of Agents and Software

This interface enables Agent to conduct interactions with other non-Agent software. First, it regards non-Agent software together as a package, and then conducts communications between non-Agent software and the FIPA Agent.

Interactions Between Agents and Human Beings

This includes how agents interact with human beings and how to operate users' information.

MAS Development Tools

The Agent-oriented software engineering approach is far from mature and the development of projects around Agent is essentially based on basic experience in practical software engineering. Many companies, in developing Multi-Agent systems, extracted the bottom service part of the systems and formed the application programmable interface abstractly to simplify the development of Agent systems or similar systems. On this basis, a large number of Agent system development tools emerged.

Generally, Agent development tools always work in heterogeneous environment which is highly distributed and often run on an Internet platform and in a web-based information system. It includes Agent language, Agent development environment or platform. The requirements for Agent language should: (1) be object-oriented;

(2) have platform independence; (3) have communication capability; (4) possess security features maintenance; (5) have code manipulation. The languages VC++, Java, LISP and so on can meet these requirements. The Agent development environment should provide the functions that the agent basic software should have.

An Agent operating environment is comprised of a group of the most wanted services which are set up in advance in each Agent. Middleware includes the distributed management environment and other services, and it also provides a unified port interface. The operating system and communication system provide the basic needs of computer operation and communication. There are now many commercial Agent development environments and platforms which can be used to easily create Agent based practical systems, such as Agent Builder (Reticular Systems Inc.), Microsoft Agent (Microsoft Inc.), Aglet (IBM), Cable (Logica Corporation), Cybele (Intelligent Automation Inc.), Agent Talk (NTT Inc.), and JATLite (Stanford University). Next, a few simple Agent development tools will be introduced.

Agent Builder

Agent Builder is an integrated software development environment, which can be used to create Agent based applications. Agent Builder includes operating systems and tool kits in two parts.

Tools in the tool kits can be used to: a) define behaviors of a single Agent. b) design and develop interactive and collaborative agent networks. c) manage the Agent based software development process and d) debug and test Agent based software systems.

When running a system, there is an Agent engine, providing operating environment for Agent software. Agents constructed by Agent Builder communicate by the use of the language KQML (Knowledge Query and Manipulation Language). In addition, Agent Builder also allows developers to define new interactive communication orders when needed.

ADK

ADK (Agent Development Kit) mainly contains the two parts: AFC (Agent Foundation Classes) and ARE (Agent Runtime Environment).

AFC is an advanced API, supporting developers' design and constructing components; ARE is a virtual server, supporting thousands of Agents running at the same time.

Moreover, there is a VAD (Visual Agent Designer) for designing an Agent. VAD can be used to design and develop an Agent quickly and easily.

LPA Agent ToolKit

LPA Agent ToolKit, based on the logic programming language Prolog, allows the designing Agent-oriented programs using logic based methods, and has been able to make full use of AI technology. It provides the necessary components and can implement any Agent models, and such flexibility is especially suitable for scientific purposes. The ToolKit also provides an Agent skeleton which can be used to rapidly generate users' own Agents.

JATLite

JATLite is the first Java based typical agent development platform, its main feature being that it implements the Agent language KQML to execute management functions. JATLite is a tool that enables users to create new Agent systems quickly; it uses an Agent Communication Language (ACL) as the underlying language of communication, such as KQML. JATLite provides a method that enables each Agent to communicate without knowing the IP address of other Agents, but an Agent information router is used to ensure the information transmission among agent., The

advantage of this approach is particularly obvious in circumstances where the addresses of agents change during the computing process. JATLite provides communication functions for Agent systems without additional restrictions in other areas. It can neither know the Agents' internal structure and composition, nor whether is this Agent mobile.

Problem Solving and Interaction Mechanisms in MAS

Problem-solving in MAS is a natural result of interactions between various automatic Agents. The cooperation types in MAS include task-sharing, result-sharing and so on. Task-sharing means that Agents divide a task into several subtasks and each Agent is responsible for resolving each subtask with as little communication as possible. Result-sharing means that Agents cooperate with each other through sharing their results.

1. **Problem-solving with common objectives.** When accomplishing common tasks with a group of Agents is more efficient than accomplishing common tasks individually, then these multiple Agents should cooperate with each other to solve problems.
2. **Problem-solving with different objectives.** When every Agent with different objectives cannot accomplish its own task and the efficiency of outsourcing these tasks is substantially higher than accomplishing them by itself, then these tasks should be outsourced to other Agents.
3. **Group Model.** This model means that Agents with common objectives should try to resolve problems as a team. Based on the theories of collective planning and common intentions, Tambe put forward the group model STEAM, which is comprised of group states and group operators. Group states describe the components of a group while group operators can be regarded as joint commitments on a collective behavior.
4. **Alliance Model.** This model is based on game theory. By dividing a group of individual Agents into several Agent alliances, each alliance can be viewed as a special Agent group with certain tasks. Basic procedures to form alliances are as follows: first, determine the structure of alliances; second, determine alliance values; third, allocate alliance values among group members. By repeatedly going through these procedures, stable alliances can be worked out. A stable alliance should meet the following requirements: a) individual rationality. The efficiency of being in an alliance should be higher than working alone. b) group rationality. Alliance values should be thoroughly allocated to members of alliances. c) alliance rationality. Sub alliances in an alliance shouldn't be able to break away from their parent alliance to form new alliances.
5. **Negotiation.** Collective opinions should be formed among Agents through certain negotiating strategies. Each individual should try to enhance the influence of its own opinions and thus get the best result.

MAS mainly explores the issue of how a group of physically or logically dispersed Agents coordinate their respective knowledge, objectives plans and intentions with a view to taking collective measures to solve problems. Every individual Agent in MAS is autonomous and is capable of fulfilling certain tasks. However, it is also equipped with open interfaces which serve as accesses for other Agents to communicate with it. In this way, an individual Agent can accomplish multiple tasks each time. Social interactions between Agents in MAS include multi-Agent cooperation, multi-Agent coordination, and multi-Agent consultation.

MA coordination means that varieties of agents with different objectives coordinate their actions and distribute resources appropriately so that both individual and collective objectives are achieved. MA collaboration means that a group of Agents cooperate with each other to achieve common goals. MA consultation means that a group of Agents negotiate with each other to resolve a conflict about a course of action. To explore MA coordination and collaboration with game theory is a popular research objective.

The methods for MA coordination, collaboration and consultation include:

1. No communication collaboration and coordination: the method is based on traditional game theory.
2. Communication collaboration with central control: the method is somewhat similar to that of the operating system.
3. Consultation: the typical examples of this method include Agent consultation based on Nash's cooperative policy and consultation theory based on Rubenstein theory (Rosenschein & Zlotkin). Consultation is one of the methods to resolve conflicts between agents. The evaluating criteria for consulting mechanism and policies are as follows:
    a. Symmetrical distribution: all Agents should be on an equal footing and no privileges should be given to any Agent whatsoever.
    b. Effective: the results of consultation should not be in conflict with other results.
    c. Stability: results of consultation should be fair for all Agents.
    d. Simplicity: the cost and computation of consultation should be very low.

Applications of MAS

Researchers in different areas have paid great attention to MAS, for the systems implemented by MA technologies have obvious advantages. MAS have been applied to various practical fields, such as manufacturing, telecommunications, information management systems, business, intelligent control, medical diagnosis and education.

The Application in Manufacturing

The modern manufacturing system is a highly decentralized system, consisting of a wide range of manufacturing resources, such as standardized or non-standardized, autonomous and semi-autonomous processing equipment, material-ferrying equipment, robots and so on. It is therefore the ideal application for MAS.

The application of MAS in manufacturing industries can manifest itself in the following aspects.

1. Production and information management. Gao Guojun and his colleagues in ShangHai Jiaotong University have developed a reconfigurable enterprise information system by dint of MA technologies. This system has substantially enhanced the flexibility of companies. Further field, Swaminathan has modeled for the dynamic behaviors of supply chain by using MAS. Mehra and Nissen have developed Agent-based flexible supply-chain management system by means of Agent Development Environment (ADE) and Java. Shaw has put forward the idea of using Agent to conduct manufacturing scheduling and control. He has also pointed out that manufacturing units can subcontract tasks to other manufacturing units through tendering and bidding.
2. Manufacturing process control. A distributed control system in a discrete manufacturing environment is a control system for flexible manufacturing. Liao Qiang from Tsinghua University has provided a field-bus-based solution to workshop dynamic scheduling. In this solution, different Agents are connected

by field buses and a scheduling method is produced through coordination. Liu Jing from Zhejiang University has developed a MAS-based manufacturing process control system for steel-making.

3. Robot technology. Lane and colleagues developed single-robot MAS, with real-time blackboard Agent as the core technology. Chen Renji and his colleagues from Shanghai Jiaotong University have developed a MAS-based DAMAS (Distributed Agent-based Multi-robot Assembly System).
4. Parallel intelligent design and CAPP design. Shi Chun and his colleagues from Northwestern Polytechnic University have applied MAS technology to a parallel product design system. The application of MAS technology has made this system more stable and more flexible. Yan Li and her colleagues from Chinese Liaoning Institute of Technology have applied MAS technology to cooperative design and some key technologies have been realized along the way. Zhao Shiguang from ShangHai Jiaotong University has developed a MAS-based CAPP system. Compared with a traditional CAPP system, it is more stable, more efficient and more feasible.

Commercial Applications

Commercial applications mainly include Agent systems for information management, e-commerce and business process management.

Medical Applications

In the health care industry, new applications are constantly emerging, so the application of Agent technology is inevitable. The first two typical applications concern patient monitoring and health checking.

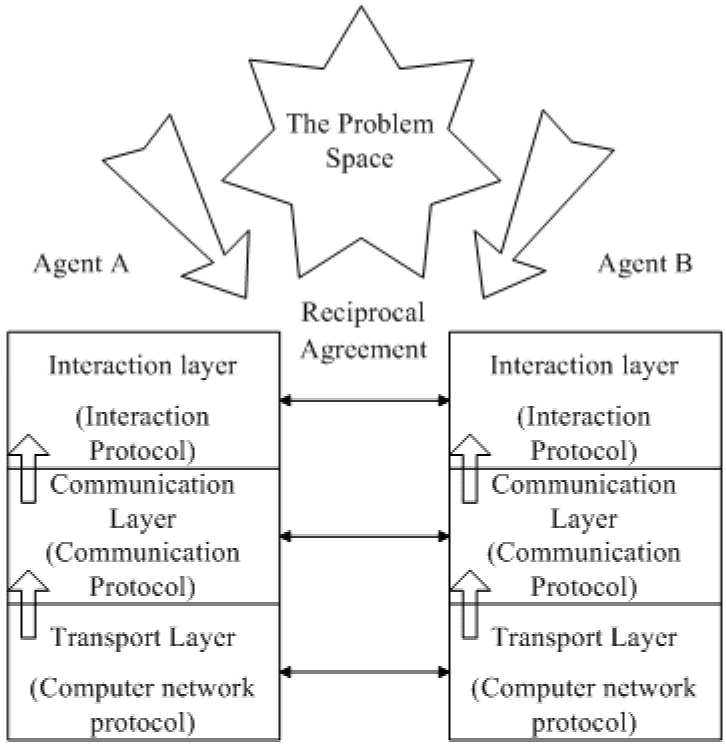
COMMUNICATION AND INTERACTION PROTOCOL IN MAS

Model

As a major way for Agents to communicate with the outside environment, the interactive activity between Agents is a clear indication of social ability. This is also an important difference between Agent and traditional AI. A sound and perfect interactive mechanism serves as the basis for MA coordination, collaboration and consultation and the precondition for sound relationship between Agents in MAS. The interaction mentioned here is a very broad concept. It refers to all kinds of interactive activities. In this sense, all the coordination, collaboration, communication and arbitration between Agents is classified into this category. The interactive mechanism of Agent mainly deals with the following questions: 'Why interact?', 'Who to interact with' and 'How to interact?'. If there is a certain environment in which Agents can work and interact effectively, then this environment is the framework. Since computer network protocols follow a hierarchical structure and the message-exchanging process follows a hierarchical structure, this framework is also hierarchical. The whole interactive behavior is divided into three layers, namely, the transmission layer, the communication layer and the interaction layer. Interactive activities taking place in lower layers provide services to interactive activities taking place in upper layers. Lower layers are the basis for the upper layers (Gao & Chen, 2001). Figure 3 gives a description of the communication and interaction models.

The first layer (bottom layer) is the transmission layer, also known as computer network protocol layer. This layer is application-oriented and is responsible for translating messages from the communication layer into certain network protocols. It also plays an important role in ensuring the realization of interactive activities. The network

Figure 3. Structure hierarchical model of communication and interaction in MAS



protocols mentioned here refer to a wide range of familiar names, such as TCP/IP, HTTP, IIOP, and so on. The theories and applications of this layer are not an emphasis of this book.

The second layer is the communication layer, also known as the communication protocol layer. This layer plays a pivotal role in ensuring that Agents can exchange and understand messages. The majority of these messages have a strong purpose, such as orders, commitments, suggestions, and objections and so on. In current research, this layer is mainly based on speech act theory and widely accepted communication languages include KQML and ACL.

The third layer is the interaction layer. This layer is mainly responsible for ensuring that agents can exchange messages with certain structures. To this end, a series of coordination, consultation and collaboration activities must be conducted. The interaction layer is one of the most important theories for the interactive mechanism. Typical interactive layers include blackboard structure, contract net protocol, voting protocol, and auction protocol.

Communication

Communication Objectives

When making predictions for changes in the surrounding environment, one Agent should keep in mind that the activities of other Agents are not under its control. In order to make accurate predictions for changes in the environment, to enhance its own ability and to reach its objectives, Agent must communicate with other Agents. The objective of Agent communication is to improve itself so that higher objectives are achieved. Through communication, Agents can better coordinate their behaviors and actions and the whole system

will be more harmonious. Communication is a clear reflection of Agent's social ability, but it is not an essential quality for an intelligent Agent. Communication is a process for the realization of Agents' objectives. In fact, communication and interaction are the exchange of mental states between Agents. The main objectives of communication are as follows:

1. **Prediction.** Through communication, Agent A can predict Agent B's behaviors after getting a clear understanding of B's beliefs and wishes.
2. **Control.** Agent A can deliver its plans to Agent B. and if accepted by Agent B, Agent A will actually put Agent B under its control.
3. **Belief.** Communication can promote knowledge-sharing between Agents. In this way, every individual Agent's vision and database will be substantially enhanced and its lack of reasoning ability is not going to stand in the way of its effectiveness.
4. **Desire.** Communication can give an individual Agent a clear understanding of other Agents' desires. In this way, this Agent can make comprehensive predictions for other Agents' behaviors and MA collaboration can be achieved.
5. **Intention.** By incorporating other Agents' intentions into its own, Agent can conduct planning collaboration (result-sharing collaboration).

To sum up, the objective of communication is to gather knowledge. This is the precondition for MA cooperation.

Communication Content

According to the Agent's mental states, communication content is divided into two basic types, namely beliefs and aspirations.

Belief Transmission

Under normal circumstances, belief transmission between Agents is a win-win activity, because when the sending party delivers a belief to the recipient, it actually enlarges Agent A's observing ability and scope. These will help and support Agent A. In some situations when the incoming belief is in logical conflict with the existing belief, the whole belief is regarded as a negotiating process. Belief transmission exerts an invisible and indirect influence on Agent's behavior, but it plays a pivotal role in the evolution of the MA group.

Intention Transmission

Intention transmission can exert direct influence on Agent behaviors and also plays an important role in the coordination of Agent's behaviors. Since intention has the same expressing method as belief, it is transmitted just like belief. When an intention reaches its target place, Agent A should decide whether to treat the intention as its own intention and decide what moves to take.

Message Types for Communications

Communications between Agents with different levels of ability are essential to the normal function of the whole system, so it is necessary to define Agent's ability and decide message types. In order for Agents to be interested in each other, dialogues must be carried out between Agents. They can play either a positive role or negative role in the dialogues. but no matter what type of role they choose to play, one thing is essential - the ability to receive. There are two basic message types: declaration and request.

When an Agent interacts with another equal Agent, it can play both a positive role and negative role in the dialogue. However, it should also make declarations and requests.

Communication Process

The communication process of two Agents in the real world is as follows:

1. The sending party translates its own thoughts into the communication language.
2. The recipient adds a language format to the communication carrier, such as voice, text and images.
3. The communication carrier reaches the recipient.
4. The recipient extracts the language codes from the carriers.
5. The recipient translates the language codes into ideas and familiarizes itself with the thoughts of the sending party.

There are two objects involved in the communication process: the Agents and the communication carriers. The moment the communication carriers leave the sending party, they are not under the control of the sending party.

Message-Sending Process

First Agent *A* should make sure that there is an Agent *B* and what kind of language and behavior models it can understand. Information *I** is ready to be sent.

1. Create the communication carrier by the formula of *C*=M(*I**). *C* mentioned here could be a language entity or a group of behavior plans.
2. If *C* is a language entity, then set the receiving environment of the recipient *B* as Rcv(*B*).
3. If *C* is a group of behavior plans, then Agent should act in accord with it.
4. Labeled *I** as "Sent" in Agent *A* information sheet.

Message-Receiving Process

Let us suppose that Agent *A* is available and model *M*, and receiving environment Rcv(*B*).

1. To make sure whether there is any new communication carrier *C* in Rcv(*B*).
2. If there are new carriers, then choose appropriate model *M*.
3. Add Send(*A*,*T*,*I**) in observing beliefs.
4. Decide whether add *I** ingot observing beliefs according to Trust(*B*,*A*).

Interaction Protocol

Communication protocols of Agent ensure that all communicating parties can exchange and understand messages, but in order for agents to interact with each other, a series of organized message-exchanging processes must be conducted. These message-exchanging processes are called 'dialogues'. Interactive protocols are generalizations for those stable dialogues.

Interactive protocols are clear reflections for the objectives and rules of Agent interactions. They are also closely related to the internal reasoning mechanism of Agents. Therefore, research into interactive protocols is crucial for MAS research. Judging from existing interactive protocols, they are all reflections of many kinds of human interactive activities. They are also combinations of scientific findings from a wide range of fields, such as sociology, histology, decision theory, artificial intelligence and so on (Shi, 2000).

The Classification of Interactive Protocols

In order to accomplish a certain task, many rounds of message-exchanging must take place. In response to different tasks, the dialogues between agents will follow fixed patterns. Message-sending and message-receiving will follow fixed sequences. This typical pattern of message-exchanging is the interactive protocol. FIPA is comprised of different and important interactive protocols, such as accept-proposal protocols, agree protocols, cancel protocols, CFP (call for proposal) protocols, confirm protocols, disconfirm protocols, failure

protocols, inform protocols, propose protocols, not-understood protocols, proxy protocols, request protocols, refuse protocols, reject-proposal protocols, and subscribe protocols.

Interactive protocols are classified by a wide range of standards, two of which are very important. The first classification standard is the interactive objective. By this standard, interactive protocols are classified into the following three categories, coordination-based, collaboration-based and consultation-based interactive protocols. Some books and papers even refer to these three kinds of interactive protocols as simply coordination, collaboration and consultation protocols. Generally speaking, coordination protocols have very broad scope. The objective of this type of interactive protocol is to reach uniformity between MA behaviors. We can equate coordination interactive protocols to interactive protocols. Collaboration protocols mean that interactive parties are somewhat similar to each other or have temporary common interests. In order to reach common goals, they have to collaborate with each other and interact with each other. Typical collaboration protocols include the contract net protocol and blackboard protocol. The meaning of consultation protocols is the opposite to that of collaboration protocols. This type of interactive protocol is competitive and selfish. The objective of this type is to seek the best possible results for the Agent alone. Typical consultation protocols include voting mechanism, auction mechanism, negotiating mechanism and debating mechanism.

The second standard for classification is the longevity of the interactive effect. By this standard, interactive protocols are classified into the following three categories, long-term, mid-term and short-term protocols. Long-term protocols define interactive protocols for Agents over a long period of time. The commitment of Agent to the role it plays actually puts restrictions on the interactive model and content. In this way, a long-term interactive protocol will be established between Agents. By contrast, short-term protocols only define interactive rules for Agents in a certain task. These rules may be effective only once, such as in the contract net protocol. As a matter of fact, there are no clear distinctions between these classification standards. For special purposes, the designer of Agents can adopt integrated interactive protocols. Irrespective of the kind of protocol that is used, the ultimate goal is to ensure that the dialogues are conducted in an organized way and the objectives are reached.

The Description of Interactive Protocols

From the interactive objectives and interactive processes, we can provide an overall description for all kinds of interactive protocols.

Let us suppose that A is a set consisting of all Agents; S stands for all possible states for Agents; B stands for all possible interactive behaviors of Agents. The power set of a random set Q is $\wp(Q)$. So an interactive protocol InP (Information Navigation Protocol) is defined as a 6-member array:

$$Inp=\langle Ag,St;Ob,R,Bh,Act\rangle$$

In this array, $Ag\in\wp(A)$ stands for objects involved in this protocol and it includes all Agent entities with an interest relationship, such as contractors and customers in the contract net protocol.

$St\in\wp(S)$ stands for all possible interactive states in the process of interaction. These states are closely related with the interactive behaviors of Agent. Let us take contract net protocol as an example. When bidding documents are delivered by customers to all potential contractors, customers have a new state. After receiving the bidding documents from contractors, customers have another new state. Finally, when the problem is solved, customers will have the final state.

$Ob\in\wp(S)$ represents the final state of a certain interactive protocol. This final state is a clear reflection of Agent's rational pursuit and it

often manifests itself in the form of some kind of evaluating index. In addition, this state can either be predefined or only be clear in the interactive process.

$R \subseteq St \times St$ represents the direct relationship between interactive states. This relationship, coupled with Agents' behavioral rules, is conducive to getting a clear route from the initial state to the final state.

$Bh \in \wp(B)$ represents all acceptable behaviors in the interactive process. These behaviors are further divided into two types, positive behaviors and negative behaviors.

$Act: R \times St \rightarrow Bh$ denotes, under the function of states relations, that the change of the interactive states is relative to a certain interactive behavior. On the one hand, it shows that the change of Agent's interactive states is always related to a certain behavior implemented by the Agent. For example, in Contract Network Protocol, if the status of the contractor is changed, it indicates that the contractor has received a certain bid, has issued a bid, or a contract has been completed. On the other hand, any one effective act implemented by the Agents must also be related to the corresponding interactive states according to the states relationships. In the Contract Network Protocol, under the state that the conductor received the bid returned by the potential contractors, the conductor's effective behaviors should be to evaluate and select the designated objects, and then to send the message to confirm the factual contractor.

The process mentioned above gives a general description of interactive protocols, but in practical applications, special emphasis must be placed on the following two conditions essential to smooth running of protocols. The first is the working environment; for example, contract net protocol is based on the assumption that Agents are cooperative, not hostile toward each other, and network resources are adequate. The second is that extra norms and regulations must be put forward in accordance with the specific operating environment.

In order to avoid deadlock of interactive protocols in a real environment, a verification process should be added to the design and analysis of interactive protocols. Lacey & Deloach (2000) give detailed information about this verification process.

COOPERATION AND BEHAVIOR COORDINATION

Cooperation

Cooperation is one of the basic concepts in MAS (Ferber, 1999). Cooperation is the technology for distributing tasks, information and resources among Agents working together. Cooperation actually addresses the issue of 'who does what?'

Functions of Cooperation

There are three main functions for cooperation, namely, enhancing the system's survival ability, improving the system's performance and resolving conflicts. There are three corresponding indicators, namely, the survival indicator, performance indicator and conflict indicator.

Enhancement for the System's Survival Ability

The survival indicator gives us a clear reflection of the capabilities of the system to function normally despite damages that have been done to the system. Even simple cooperative mechanisms can boost Agents' abilities to survive. There are two kinds of survival, individual survival and collective survival. Agents' individual survival ability is defined in two ways. The first one is the Agent's survival possibility in a certain environment. The other is the energy efficiency factor, which is the ratio between the absorbed energy and the consumed energy in the unit time. If this energy efficiency factor is larger than 'one', then both individual Agents and the Agent group can

survive. The first way is suitable for the analysis of simple Agents' behavior, but the second is more universal than the first.

Improving the Performance of the System

The performance indicator shows an Agent or a group of Agents' abilities to accomplish tasks assigned by observers. These observers could belong to this group of Agents, or might not. Although there are many kinds of performance indicator, all tend to display the basic feature of collectivism. Moreover, we can draw parallels between performance indicators and energy efficiency indicators.

Generally speaking, we can increase efficiency by raising the number of Agents, but if there is no cooperation between Agents, boosting efficiency will be impossible. When there is cooperation, more Agents can work in coordination to achieve common goals and improve efficiency. Therefore, a group of Agents cooperating with each other will be more efficient than a group of isolated Agents.

Collaboration plays an important role in the improvement of a system's performance both quantitatively and qualitatively. Generally speaking, collaboration enhances the capability of the individual Agent. Qualitatively speaking, collaboration can enable a group of Agents to accomplish tasks that are hardly possible for individual Agents to complete. Without collaboration, a task that cannot be accomplished is a 'mission impossible' for a group of agents. The basic function of collaboration is therefore to qualitatively improve the system's performance. This is particularly important, since quantitative improvement sometimes is transformed into qualitative improvement and vice versa. An auxiliary Agent that is being assigned a special role has the potential to substantially enhance the overall performance of a system. It is somewhat similar to a catalyst in chemistry. As an example, the existence of traffic police will generally significantly improve the flow of car traffic at a crossroads. This is the result of traffic police having the authority to control the flow of cars. Likewise, collaboration can improve a system's performance quantitatively and qualitatively; the result of the collaboration can also be quantitative and qualitative.

Conflict Resolution

The conflict indicator easily reflects the number of Agents in a conflicting environment and the number of entities operating resources simultaneously. Due to Agents' autonomy, conflicts between Agents are very common. The origins of all these conflicts are the limited resources. For example, if two agents all want access to resources that cannot be shared by more than one Agent, conflicts will take place, and since the resources that are available are often limited, conflict-resolution mechanisms must be adopted. Although total conflict or total cooperation is possible, the most common situation will be the coexistence of conflicts and cooperation, and most complicated systems are combinations of a conflict environment and a cooperation environment. Cooperation plays a role in improving the system's performance and conflict makes it possible for us to choose Agents.

Cooperation Theory

Since knowledge is widely distributed, it is never possible for one Agent to accomplish higher goals all by itself. The realization of these objectives involves a great deal of teamwork, which is the basic theory for cooperation between MAS and Agents. In an open, dynamic and MA environment, Agents with different goals must coordinate their objectives and resources. There are three main theories for the cooperation between Agents.

1. **Shared intentions.** The precondition of this theory is that all Agents should share the same objectives. The most important finding of this theory is that when an Agent adopts a thought that is not shared by other

Agents of the team, it has the obligation to notify other members of the team.

2. **Shared plans.** If a certain Agent is not capable of resolving a problem, it may establish a contractual relationship with other Agents and outsource this problem with them. To this end, these other Agents must have the following features: must be able resolve the problem, and to resolve it at a specific time.
3. **Planned team activities.** When a certain Agent wants to accomplish certain goals but is not capable of doing so, the need to form a group is manifested. This Agent will communicate with potential partners and notify them of common objectives.

Collaborative Approaches

Collaborative approaches refer to methods to realize MA collaboration. Some experts classified collaborative approaches into the following 6 categories: grouping and teamwork, communication, specialization, collaboration through sharing tasks and resources, coordination of behaviors, and conflict resolution through consultation and arbitration.

Grouping and Teamwork

Grouping is the simplest collaborative approach. This method is somewhat similar to the world of gregarious animals. The reasons why animals want to be in a group are diverse, such as for safety, for food and for breeding. Grouping resolves the problem of navigating. That is to say, if an Agent or Agents in a group have decided where to go, other Agents are relieved of this decision and just follow the lead of those pioneering Agents.

Teamwork means that the increase of individual entities in a system can boost the system's performance and reliability. If a system only has one Agent, then the task cannot be accomplished if this Agent cannot function properly. In this sense, the system's reliability will be substantially boosted by the existence of multiple Agents. If there is no glitch in the system, then its performance will also be significantly enhanced.

Communications

Communication is not only essential to MA collaboration. It is also one of the basic methods for the distribution of tasks and coordination of behaviors. Communication can establish links between groups of unrelated Agents. In this way, Agents can get a clear understanding of other Agents' information and knowledge, and Agents' understanding ability will be substantially enhanced. In a cognitive system, communication is realized by message-exchanging; in a reactive system, communication is realized by the transmission of signals in the environment.

Specialization

The objective of specialization is to familiarize Agents with tasks and to enable them to accomplish tasks in a quick and effective way. Developing Agents that are suitable for all kinds of tasks is not impossible; if a particular type of Agent can accomplish a certain task efficiently, then their internal structure and behavioral characteristics make it almost impossible for them to accomplish other tasks. A powerful Agent can gradually become more specialized in the process of accomplishing a temporary task. This specialization is conducive to the functioning of the whole system because it enhances the system's ability to effectively address similar problems.

Collaboration Through Sharing Tasks and Resources

There are numerous ways to distribute tasks in a system. In a cognitive system, tasks are distributed by a supply-and-demand mechanism, which is divided into two categories, the centralized distribution mechanism and the distributed distribution mechanism. The centralized distribution

mechanism means that a central Agent has controlled and distributed supply and demand. The distributed distribution mechanism means that Agents can provide supplies and make demands simultaneously and no centralized mechanism is needed. There are correspondingly two types of technologies. The first is based on the cognitive network, and the other is based on the market mechanism, and tendering and bidding protocols are needed.

The Coordination of Behaviors

In MAS, it is necessary to run an extra task, namely, the coordination task. Although these tasks do not have a direct influence, they play an important role in ensuring that tasks can be better accomplished. For MAS, coordination tasks are essential because manufacturing tasks depend on coordination tasks.

Conflict Resolution Through Consultation and Arbitration

In MAS, consultation and arbitration are often adopted to resolve conflicts so that the system's overall performance will not be diminished. Arbitration means that behavioral rules must be defined to rein in Agents so that conflicts are avoided. In this way, both individual entities and the whole system are protected. In human societies, people's behaviors are restricted by laws and regulations, but in MAS, these restrictions are frequently defined in Agent itself.

The consultation technologies for cognitive MAS and reactive MAS are totally different. For cognitive MAS, consultation technology is very important because when Agents are in conflict with each other for objectives and resources, they tend to resolve their conflicts by negotiation. For reactive MAS, consultation technology is no longer important because the behaviors of reactive Agents are fixed, and there is no possibility that Agents will overstep these restrictions.

Collaboration Process

When a certain Agent believes that good can come from collaboration, the desire to cooperate with each other will manifest. When a large group of Agents found that they can make bigger achievements through collaboration, they will form an Agent coalition and take concerted actions. While there are some differences in the reasons for collaboration, they often follow this rule: Agents should compete for positions and roles and the assignment of positions and roles should be carried out in a fair manner. By putting planning, competition, restrictions and cooperation into a single framework, we can divide the cooperation process into the following sub-process:

1. Evaluate requirements and establish objectives.
2. Carry out cooperative planning and define cooperative structure.
3 Find cooperative partners.
4. Choose cooperative policies.
5. Realize the objectives.
6. Evaluate the results.

The process of finding cooperative partners is actually the process of choosing cooperative policies. In order to improve the efficiency of finding cooperative partners, it is necessary to optimize the cooperative structure and eliminate those cooperative policies that stand no chance of reaching a final destination. The process of finding cooperative partners can be described as follows:

1. The coordinator of the cooperation should first disclose the positions and roles that are available for competition.
2. Competing Agents should bid for those positions and roles in accordance with their own capabilities, and put out relevant information, such as the amount of time that is

needed, resources that are necessary and so on.

3. The most appropriate Agents will be picked as cooperative partners in accordance with the available information. As long as commitments are made by cooperative Agents, they have to try their best to honor those commitments. If there is any Agent that cannot fulfill its obligations, the cooperative relationship will be severed immediately.

In order to accomplish common objectives, Agents' cooperative behaviors should meet the following criteria:

1. Agents should be able to respond to each other.
2. Agents should make reasonable commitments to concerted actions with other Agents.
3. Each Agent should be able to support concerted actions.
4. Cooperative Agents should be able to meet certain environmental requirements.

Behavior Coordination

Coordination and cooperation are key issues that are studied in relation to MAS. Coordination refers to the interactive characteristics exhibited when groups of Intelligent Agents complete collective activities, including adaptation to the environment. Effective coordination is the key point that the autonomous Agent achieves in MAS. Because of the interdependent actions of Agent in MAS, the multi Agent intention exists. Once conflict or cooperation occurs, behavior coordination is needed.

The Definition of Behavior Coordination

There are many definitions of behavior coordination, but Thomas W. Malone's is representative. He described behavior coordination as an assistant act collection, and these acts are implemented in MAS. Also Agent could not accomplish the task all by himself. Behavior coordination could be defined as management between the interdependent relationships. There are many examples of behavior coordination; for example, if two people intend to leave a room, and walk toward to the same door, which is too narrow to allow more than one person to pass, one person must stop to let the other person pass first.

Behavior coordination is the guaranteed primary means for Agent to coordinate in MAS. Through the coordination of Agents in MAS, the overall performance of the MAS is improved, and conflict within the system is reduced. The main reasons for MAS to carry out behavior coordination are as follows. Firstly, the inputting information that the Agent needs is provided only by other Agents. Secondly, many resources are limited, so when available resources are limited and shared, behavior coordination must be carried out. If the sharing resources reduce, then behavior coordination will become even more important. Thirdly, it is necessary to reduce costs. Unwanted behavior and redundancy can be eliminated through behavior coordination, thereby reducing costs. Finally, the objectives of the Agent are compatible and interdependent, but not independent, so the Agent has to consider the behavior of other Agents to avoid disturbing other Agents. If possible, it will also help other Agents to achieve their objectives.

The Problems Needing to be Solved by Behavior Coordination

Behavior coordination has brought about new problems, such as what to be coordinated with, how to detect and remove potential conflicts, and what method should be selected to carry out the coordination.

Table 1. Eigenvalues of different cooperation forms

	Synchronous	Planning	Reactive	Rules
Fast	Very Good	Bad	Excellent	Good
Adaptability	Very Poor	Bad	Excellent	Good
Predictability	Bad	Excellent	Bad	General
Concentration and Distribution	Irrelevant	Irrelevant	Irrelevant	Concentration
Communication	Message	Message	Incentive/Tag	Irrelevant
Freedom	Very Limited	Limited	Large	Relatively Limited
Quality of Cooperation	Relatively Good	Excellent	Relatively Good	Relatively Good
Conflict Resolution	Good	Good	Bad	Good
Number of Agents	Large	Small	Very Large	Large
Number of Exchanged Data	General	Large	Small	Small
Degree of Expression	Limited	Countless	Limited	Limited
Difficulty of Realization	General	Big	Small	General
Heterogeneous	Low	Very Low	High	General
Universality	Bad	General	General	General

Determining the Coordinating Target

In some cases, it is simple for an Agent to determine what to coordinate with. In air traffic control systems, mobile robots, and anti-collision systems, for example, the Agent needs to coordinate with the nearby aircraft, robot or vehicle. Generally speaking, the targets needing to coordinate have normally been determined by the problems themselves.

The Interaction Between Behaviors

It is not enough to determine the coordinating objects, as interactions between Agents in MAS must also be managed. MAS which is formed by only two Agents does not need to consider the interaction between Agents. However, MAS which is formed by three or more components of Agent have to consider the interaction between Agents. Generally speaking, if there is coordination, any acts of an Agent are dependent on all other Agent acts. When limited resources result in a conflict, the Agent which has more limitations could change its behavior hardly. On the contrary, the Agent which has little limitations could take other Agents' behavior into account, and to respond quickly.

The Relationships Between Behaviors

The relationships between behaviors are divided into two categories, positive and negative relationships, by F. von Martial (Ferber, 1999). Negative relations are those that prevent the implementation of several acts at the same time, and are generally caused by limited resources and incompatible goals. Positive relations benefit Agent. Positive relations include equal, preferential and inclusive relationship. The equal relationship means that some certain acts are not linked to a specific Agent and implemented by other Agents. Preferential relationship means the possibility that Agent preferred to implement another act while it has a certain act to be executed. Inclusive relationship means that act of Agent ***A*** result in act of Agent ***B***.

The Forms of Behavior Coordination

There are four main forms of behavior coordination: synchronous coordination, planning coordi-

nation, coordination of response, and formal-style coordination. Each form can have several variations, such as shown in Table 1.

Synchronous Coordination

Synchronous coordination is a basic form of coordination, which is also a 'vulgar' form of coordination. Synchronization was originated from distributed systems. When it needs to manage multiple acts and testing operations with consistency of results, synchronization is needed.

Planning Coordination

In the field of artificial intelligence, planning coordination is a very traditional technology. The technology is divided into two stages. In the first stage, the necessary behavior to implement is determined according to the objectives and is generated. In the second stage, one of the planning is selected and implemented. As the environment may change, so in the process of implementation, it may be necessary to change the planning, which requires a dynamic re-planning method. In addition, in MAS, different planning could lead to a conflict of objectives or conflict of resources, therefore planning must be coordinated and these conflicts must be resolved to achieve the goal of different Agents. Generally speaking, planning technology can provide high-quality coordination, however it cannot be used for highly complex or unpredictable situations.

Coordination of Response

The technology of coordination of response appears much later, and is a relatively new technology. The basic starting point is to achieve a coordination mechanism based on the reaction of Agent, which is usually easier than pre-planning the behavior and interaction before implementation. In essence, this technology uses an Agent-sensing feature, and is based on the real environment, rather than based on reasoning. Although the approach of reaction coordination does not necessarily guarantee all the results of the optimization, it is capable of coordination in the evolution environment (even in an unpredictable environment).

Formal-Style Coordination

The approach of formal-style coordination is relatively rare. It is mainly used for systems needing limited coordination capacity. Its basic principle is to establish a group of rules to eliminate potential conflicts. Under normal circumstances, this method adopts rules to define what kind of behavior is 'good' to avoid conflict as far as possible.

APPLICATIONS OF AGENT IN INTELLIGENT MANUFACTURING

Many researchers have applied Agent technology to intelligent manufacturing. Boonserm et al. described a framework for facilitating the collaboration of engineering tasks, particularly process planning and analysis for globalized manufacturing activities (Boonserm et al., 2004). Liu and Yang presented an approach that utilizes a combination of information and knowledge models to support global manufacturing coordination decision-making (Liu & Yang, 2004). Jiao et al. applied the MAS paradigm for collaborative negotiation in a global manufacturing supply chain network (Jiao, 2006). Wang et al. proposed an architecture where many facilitator agents coordinate the activities of manufacturing resources in a parallel manner (Wang, 2005). Nahm and Ishikawa discussed a MAS framework for integrated product design in a computer network-oriented CE environment (Nahm & Ishikawa, 2005, 2006). It is predicted that profound changes take place in manufacturing industries by the application of MAS (Qiao & Zhu, 2001).

There are many MAS projects in intelligent manufacturing. The projects are classified into four categories: (1) enterprise integration and enterprise supply chain; (2) manufacturing process planning and production scheduling; (3) holonic

manufacturing systems; and (4) man-machine cooperation.

Enterprise Integration and Enterprise Supply Chain

Enterprise integration means that each unit of the organization will have access to the information relevant to its tasks, and will also understand how its actions will impact other parts of the organization, thereby enabling it to choose alternatives that optimize the organization's goals (Shen & Norrie, 2001). Enterprise integration emphasizes the enterprise's structure, organization, management and relationships of its internal parts. It can achieve the purpose of making the enterprise more agile and improve the enterprise's competitiveness. Nevertheless, for information exchanging, resources sharing and the transparent interoperability needed in the enterprise integration process are very difficult problems for most manufacturing enterprises.

A manufacturing supply chain is a global network of suppliers, factories, warehouses, distribution centers and retailers through which raw materials are acquired, transformed and ultimately delivered to customers. Improving the performance of supply chain management, integrating the operating systems of the whole enterprise and forming a highly unified increment supply chain system are important means to enhance the enterprise's competitiveness and profitability.

Many researchers have probed solutions for enterprise integration and the enterprise supply chain. Swaminatnan has proposed using a Multi-Agent framework for modeling supply chain dynamics (Swaminatnan, Smith & Sadeh, 1996). Mehra and Nisen studied and developed an agent based supply chain management system making use of ADE tools (Agent development environment toolkit) and Java language. The project MetaMorph, carried out at Calgary University, Canada, used a hybrid Agent-based mediator-centric architecture to integrate partners, suppliers and customers dynamically with the lead enterprise through their respective mediators within a supply chain network via the Internet and Intranets. In MetaMorph, Agents are used to represent manufacturing resources (such as machines and tools) and parts, to encapsulate existing software systems, to function as system or subsystem coordinators (mediators), and to perform one or more supply chain functions (Shen, Norrie et al., 2006). Papaioannou and Edwards have applied mobile agent technology for virtual enterprise modeling and integration. Jeff and Pan have proposed applying intelligent Agents to enterprise intelligence integration. Pancerella and Hazelton have brought forward the architecture of autonomous agents supporting agile manufacturing. Lefrancois proposed the enterprise information integration framework based on Agent technology.

The above research work has concluded that Agent technology provides a natural way to realize enterprise integration and supply chain management. By making use of agent technology to realize enterprise integration and supply chain management, an enterprise can respond more quickly to ever-changing markets. The Agent approach also enables an enterprise to greatly enhance its efficiency of information exchanging and to obtain feedback from the market quickly. Moreover, using the learning capability of Agents in Multi-Agent systems, an enterprise can configure supply chain resources rationally and constantly improve and optimize the performance of the whole enterprise system.

Manufacturing Process Planning and Production Scheduling

Agent technology has great potential in the planning and production scheduling of the manufacturing process. Manufacturing process planning is the means of selecting and sequencing manufacturing processes such that they achieve one or more goals and satisfy a set of domain constraints. Manufacturing scheduling is the process of se-

lecting among alternative plans and assigning manufacturing resources and time to the set of manufacturing processes in the plan, and optimizing the selection in accordance with a set of rules and indicators which reflect the time relationship between operations and resources. Manufacturing process planning and production scheduling is a complex issue, especially when the operating system of the modern manufacturing system is full of uncertainty and the manufacturing tasks of the system are dynamically changing, such as with unpredictable increases or reductions in missions, the shortage or over-supply of manufacturing resources, the changes of processing time for manufacturing tasks and so on. In order to be competitive, while maintaining quality and reducing investment costs, an enterprise should shorten product life cycles, accelerate a product's time into the market, increase product diversity and meet market demand in real time. These dynamic and complex uncertainties, make manufacturing process planning and production scheduling even more complex.

As manufacturing process planning and production scheduling has very important practical significance for manufacturing systems, it has been the subject of extensive research. A variety of heuristic search methods, performance matrices, Petri Nets, Genetic Algorithms, Neural Networks, Simulated Annealing and Fuzzy logic systems have been used to solve manufacturing process planning and scheduling problems. Recently, Agent technology has also been widely used to solve such problems. Shaw proposed manufacturing scheduling and control based on Agent methods, and pointed out that each manufacturing cell can send tasks to other manufacturing units as subcontracts by using the bidding mechanism. Bulter proposed ADDYMS (Architecture for Distributed Dynamic Manufacturing Scheduling) and divided scheduling into two levels. In the first level, through Agents' consultation mechanisms, manufacturing units are assigned to operations in a distributive approach; in the second level, shared manufacturing resources are distributed dynamically. A system structure for distributive real time scheduling ReDS (Real time Dynamic Scheduling) has been proposed by Hadavi. Chen and Luh conducted research on manufacturing scheduling and supply chain coordination using Agent technology. MetaMorph put forward an arbitration model based on hybrid mediator agents, making use of the bidding mechanism based on the Network Contract for manufacturing scheduling and rescheduling.

In one of his papers, Shen (Shen & Norrie, 1999) cited a total of nearly 30 international research projects which have been or are using Agent technology for manufacturing process planning and scheduling. From these studies, a conclusion has been reached that when Agent technology is applied for manufacturing planning and production scheduling, Agents always represent the variety of physical resources and logic resources, such as machine tools, materials transport equipment, robot, orders The various kinds of Agents are organized according to the manufacturing process, and connected to form a Multi-Agent manufacturing system through the network or Agent communication protocol. Finally, theories and methods of the Multi-Agent system are utilized to study Agent collaboration, consultation, scheduling, control, conflict resolution, and other behaviors in the manufacturing system.

The Multi-Agent system can lower system complexity by decomposing the tasks; also when necessary it can obtain the necessary system adaptability, robustness and retractility with relatively low cost. Moreover, the Multi-Agent system can coordinate its decision-making dynamically, self-adjust its behaviors, and give intelligent responses to changes in product and equipment failure. These important characteristics of Agent are the driving forces for the choice of Agent technology for manufacturing process planning and production scheduling.

Holonic Manufacturing Systems (HMS)

In 1967 in one of Arthur Koestler's books named 'The Ghost in Machine', he studied the structural characteristics of social organizations and living organisms, proposed the theory of open hierarchical system and observed the dichotomy of wholeness and partness in systems. He stated that 'each entity in systems has the double features of collaboration and autonomy; it is collaborative as a part of the upper entity and is autonomous as a whole for its lower entity.' In order to completely explain such characteristic of the entity in system, he coined a word Holon by himself. Holon is from Greek and combined by the world 'Holos' (Holos meaning 'whole') and the suffix 'on' (on implying 'part'). Koestler made further study on the emerge, survive and evolvement of Holon systems and came to the conclusion that Holon systems are all complex adaptive systems with the following characteristics: Autocatalytic sets and Positive Feedback; Self-similarity; Self-organization and Decoupling Mechanism.

Inspired by the concept of Holon introduced by Koestler, in the beginning of the 1900's, the international Intelligent Manufacturing Systems Research Program conducted a major project named Holonic Manufacturing system: system components of Autonomous Modules and their distributed control. The HMS consortium proposed the following concepts and presented their definitions.

Manufacturing Holon is an autonomous and cooperative building block of a manufacturing system for transforming, transporting, storing and/or validating information and physical objects'. A holon can be part of another holon.

Autonomy, which can also be called self-discipline, is the ability of the Holon to complete the assigned tasks by planning and controlling their own behaviors.

Collaboration, also known as sociality or cooperation, refers to the ability of a group of Holons to work together and carry out plans that are acceptable for every Holon.

Another important concept 'holarchy' is defined as 'a system of Holons which can cooperate to achieve a goal or objective'. Holarchy defines the basic rules of collaboration and constrains their autonomy.

A Holonic Manufacturing System (HMS) is 'a holarchy which integrates the entire range of manufacturing activities from order booking through design, production and marketing to realize the agile manufacturing enterprise'. In intelligent manufacturing systems, many entities such as raw materials, machines, products, and parts, are all autonomous and collaborative. Therefore, they all can be considered as a Holon in HMS, In an HMS, each Holon's activities are determined through cooperation with other holons, rther than being determined by a centralized mechanism.

Agent technology has been applied to implement the HMS concepts for Agent and Holon have similar concept and structure. Agent and Holon are all autonomous, collaborative and intelligent In an HMS implemented with Agent technology, intelligent agents representing the 'holons' have a physical part as well as a software part.

Man-Machine Cooperation

A human is the most intelligent part of an intelligent manufacturing system and the source of knowledge innovation. As people play a very important role in manufacturing systems, machine intelligence and human intelligence should be combined so that humans and machines can cooperate to fulfill tasks. This man-machine cooperation has received much attention in recent years.

Humans interact with other Agents through logic Agents. Through image thinking and intuitive thinking, they are always at the highest level of manufacturing systems and coordinate with other parts of the system in order to achieve the goal of the system. If logic Agents are established corresponding to the human role, then human

intelligence can be brought into full play. A sound intelligent system cannot be established based on Artificial Intelligence alone; the perfect combination of Human Intelligence and Artificial Intelligence (also known as machine intelligence) is a practical way to the ultimate realization of intelligent manufacturing.

CASE STUDY

In this section, we will present a case study about logic architecture of collaborative manufacturing resource share prototype system in net environment, and design functional module of system in details.

Background and Motivation

The formation of a global market will directly result in severe market competition, and competition pushes the whole society to develop very fast; meanwhile, it causes an extremely severe current environment for enterprises. How to respond to fast changing market requirement, and to provide high-quality and low cost products and excellent service at high speed has become a common goal and perennial topic for enterprises.

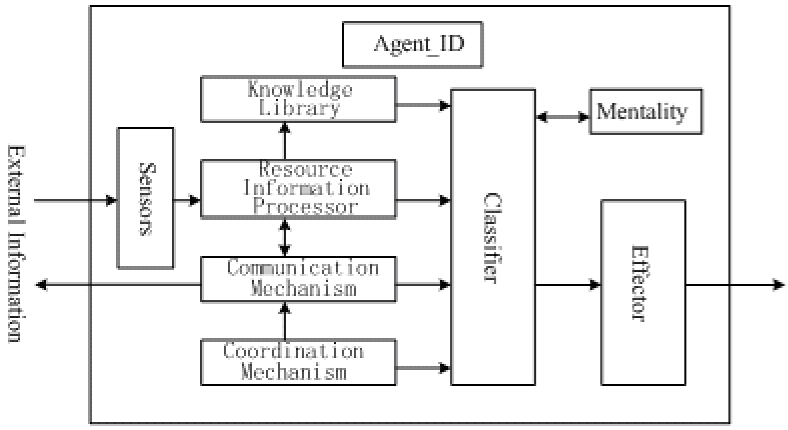
According to survey, it is shown that:

1. There is unbalanced distribution of resources in different enterprises. Many mid-scale and small-scale companies (called resource-scarce units), lack technological strength and funding; they are scattered throughout the world, so for these reasons are competitively weak.
2. Although many universities, research institutions and local productivity promotion centers (called resource-sufficient units) embrace many high-level researchers, high-level design and manufacturing equipment, advanced development tools, and strong creative and design ability, there is much wastage of talents and resources as a result of the lack of production tasks and low resources usage. Collaborative manufacturing can effectively set up complementary rules for resource-scarce units and resource-sufficient units, so as to improve resource usage, speed up production manufacturing, and enable enterprises to be more competitive.

Collaborative manufacturing resources share an original system that is one system that takes effect during the whole life cycle of the product. It is divided into two sub-systems, there are front sub-system and the back sub-system. The front sub-system is a Collaborative manufacturing resource sharing website, whereas the back sub-system is a collaborative manufacturing resource maintenance managing system. The collaborative manufacturing resource sharing website is based on Web technology, faces the enterprise Internet, and is an information platform for collaborative manufacturing resource searching and communication sharing between collaborative enterprises. The sub-system realizes two functions; one is to work as a dynamic maintenance tool for the front sub-system; another is to realize the maintenance and management of the enterprise collaborative manufacturing resource.

In this system, collaborative manufacturing resources are divided into manufacturing equipment resource, software resource, human resource, information resource, technology resource, standard resource, material resource, knowledge resource and other resources. This system is not a service system for collaborative manufacturing product design; instead, it is manufacturing resources service platform for collaborative enterprises. The central part of the system is thus the shared service and maintenance management of the collaborative manufacturing resource.

Figure 4. Structure of classification agent



Modeling

Method of Manufacturing Resource Sharing Modeling based on Multi-Agent

Category Design Agent

Category Agent (ReCA) is a nine tuple: Agent_ID, Perceptron, resource information processors, communications mechanism, coordination mechanism, the mentality, classification, knowledge base, learning mechanism and effector. The Agent_ID is used to identify the ID of the classification Agent; Perceptron is responsible for the continuing perception around the resources environment; the resource information processor classifies information resources; the communication and coordination mechanisms are responsible for the communication of the classification Agent and other Agents; mentality is the description of the state of their own classification Agent; Agent classification is responsible for physically classifying the accessed resources; the Agent knowledge base stores the various knowledge and all possible values of classification characteristics; effector is responsible for the implementation of classified results information. The structure is shown in Figure 4.

<ReCA>:: = <Agent_ID> <Perceptron> <resources information processor> <communication mechanism> <coordination mechanism> <mentality> <classifier> <Knowledge Base> <effector> <Agent_ID>:: = <Logo> <Perceptron >::=< activated conditions> <Resource Information> <Resources information processor >::=< father Resources> <subclass resources> <category of the environment> <Communication mechanism >::=< communications original language> <communications> <Coordination mechanism >::=< task coordination mechanism> <conflict adjustment mechanism> <Mentality >::=< belief> <intention> <commitment> <responsibility> <Classifier >::=< Categories> <classification based on> <classification rules> <Knowledge Base >::=< various Agent knowledge> <all the characteristics of the classification may be different attributes> <Effector >::=< Resource Information receiver>

```
< Resource category processor>
<Resources classification de-
scriptor> < Resource category>
```

The Classification Rules of the Classification Agent

In the design of the classification Agent, classification is based on the structure of the decision resources model, thus the most important thing is the decision of classified rules in the classification. Resource modeling is a organic part of network enterprise modeling tools and is closely linked to other view models; it is used for processes or activities, but also to provide support functions; it belongs to certain organizations, at the same time it is the object of information description. Therefore, the choice of classification foundation is related closely to other views and should observe the following rules:

Rule 1: the choice of classification foundation should possess all the characteristic attributes of node resources; the resource also reflects the relationship between the view of resources and other views. Assume resource attributes set as ASet, CaSet is its capacity set, DSet as foundation set, then: if $\forall D_i \in DSet \Rightarrow D_i \in CaSet$, then, CaSet⊂ASet.

Rule 2 The characteristic attribute, taken as the classification of all the resources of a node, is unique, and varies in each sub-category of resources. Assume the SRSet is the sub-class collection of resources, VSet features for the classification of all possible attributes of different values, a node have m sub-nodes, the rules for the formal description:

for $\forall D_i \in DSet \Rightarrow D_i \in ASet$,

and for arbitrary

$$\forall \operatorname{Re} s_j \in SRSet, |\operatorname{Re} s_j.\ D_i.\ VSet| = 1$$

$$\bigcap_{j=1}^{m} \operatorname{Re} s_j.\ D_i.\ VSet = \hat{o}, j = 1, \ldots, m\ .$$

The rule easily and automatically generates properties characteristics of a set of resources, to ensure that resources among the various sub-categories are independent of each other and there is no overlap

Rule 3 There are different classification foundations From the root node to all the sub-nodes, for the formal description: $D_1 \ldots D_i \ldots D_m \Rightarrow N_m$,then $\bigcap_{i=1}^{m} D_i \neq \hat{o}$.

The Classification of Resources Dynamic Model Modeling Based on Multi-Agent

For the resources in a certain environment, according to the needs of classification and rule 1 and rule 2, all the possible characteristics of classification attributes are selected. According to different resource characteristic attributes, resources are divided into sub-categories, and the operation is repeated, until all the leaf nodes turn to resource entity. From the beginning node to the leaf nodes, resources characteristics attribute are selected following rule 3.

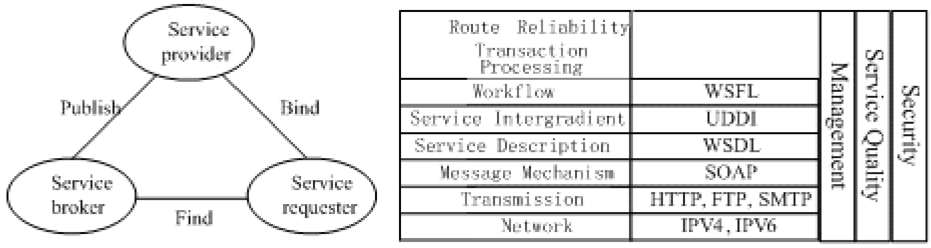
Manufacturing Resource Sharing Framework Based on Multi-Agent

Manufacturing resources sharing is to be found throughout the Internet on the basis of the effective description, integration and packaging of manufacturing resources (Huang, 2007). It is not difficult to see from the definition of the collaborative manufacturing resources chain that the nature of collaborative manufacturing resource sharing is to create a collaborative manufacturing resource chain in a network environment.

Frame Based on Service-Oriented

Figure 5 shows a service-oriented frame. Among them, Service Provider provides services, and it is components, source code or legacy system. The Service Provider publishes services to the Service Broker; requesters find the service they need through service agents, then bind to the service with the service provider. On the right of

Figure 5. Service oriented infrastructure and realization



this figure, all related technology that involves transmission is presented: messages, service description, service finding, integration, routing and working stream. Service Provider provides Web Service with the following technology:

1. XML/SOAP (Simple Object Access Protocol)(W3C, 2000) is light level protocol for information exchange in a scattered distributed environment.
2. WSDL(Web Services Description Language) (W3C, 2001) are the XML documents used to describe the network service.
3. UDDI (Universal Description, Discovery, and Integration)(UDDI.org, 2002) provides a set of public SOAP API, in order to realize a service agent. The UDDI Registry provides support for the publication and localization of the Web Service, rather like the 'yellow pages' telephone directory. The Service Requester finds the required service through the UDDI Registry, bind service provided by service provider, and implement the call.

Share Frame of Collaborative Manufacturing Resource Production Based on Multi-Agent

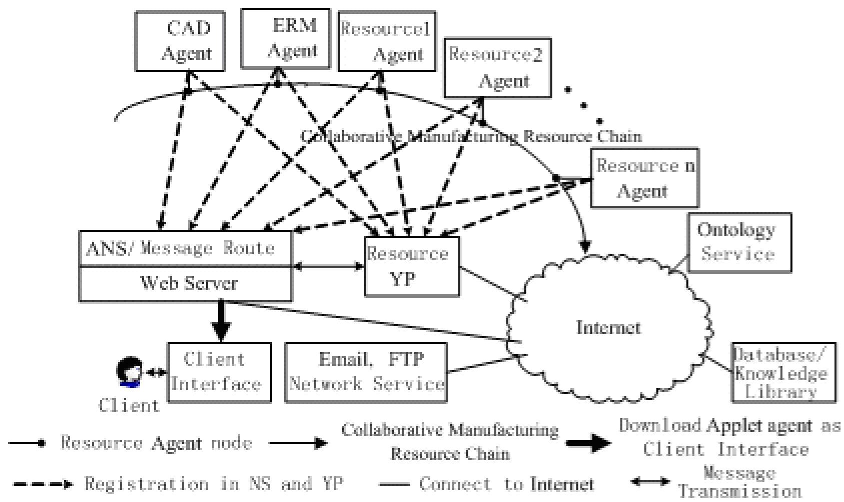
SOA provides a low-level service call frame based on PRC, but if a legacy system is used as the service provider, it is not of great value to users to work at this layer. As Figure 6 shows, the share frame of the collaborative manufacturing resource based on Agent is built by referring to SOA structure. To package Web server and form collaborative manufacturing resource set up Multi-agent system. This solution is similar to the thinking of software components, but the difference is that the share frame based on Agent copes more easily with present application systems and global sharing problems, and this is a way based on high-level message exchange and message routing.

Manufacturing resources include software and hardware resources such as operation equipments, software application system, database system, tools and knives. They are shown as Agents through agentification, with characteristics of activeness, response to the environment and intercooperation, even if sometimes the intelligent level is low or there is no intelligent ability.

In the model, there are ANS(Aggregate Naming Service), various distributed manufacturing resources packaged as Agents, user interface, Ontology service, yellow page/resource service. Remote users communicate with one or more resource Agents through an Applet Agent on the web pages. For security, the Applet Agent can only set up direct communication with the server that Applet Agent itself is on. Message Mediator therefore has to run on the WEB server. The resource service Agent is the result of agentification for an independent collaborative manufacturing resource, so it can set up direct communication with other Agents.

ANS is the central library of Agent physical addresses, and all Agents in the system have to register for ANS. ANS inside mapping relates the Agent ID with its physical address; at the same time, ANS, as a collaborative field server and router, monitors messages from each Agent,

Figure 6. Share frame of collaborative manufacturing resource production based on multi-agent



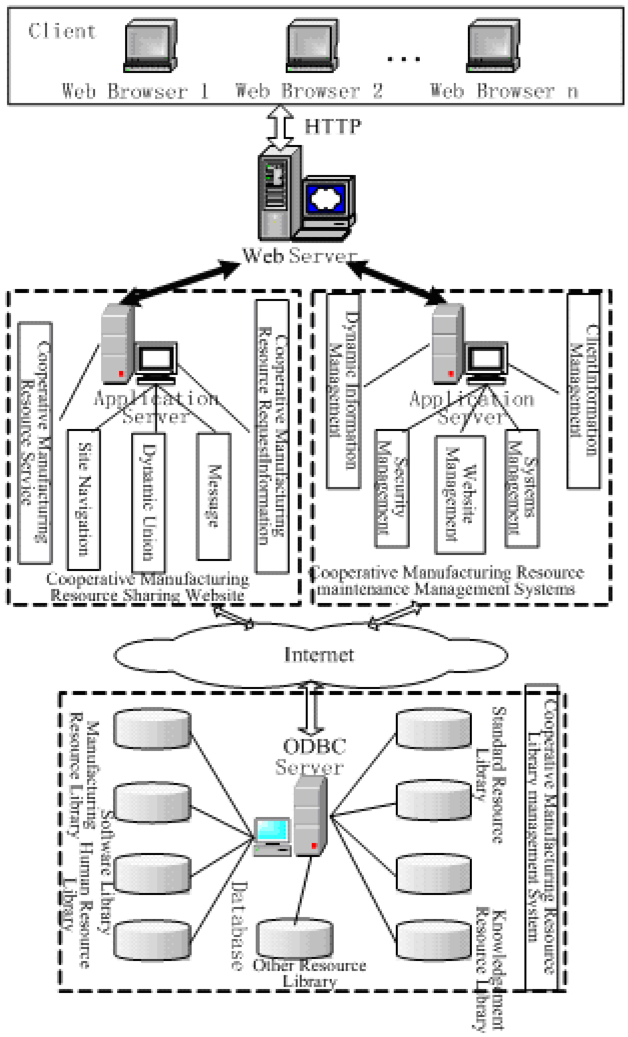
initializes information interactive ports, sets up bridges for Agents, receives messages and routes to address Agent,. Agents can communicate by Agent ID, without necessarily knowing the physical address of each Agent.

YP(Yellow Page) receives the registration and search behaviors of Agent entities, records other Agents' ability and responds to Agents' search behaviors. The user demo links users and other Agents in the system, to make operation run more smoothly. The ontology service explains the meaning of each word in the system, in order to communicate very accurately. Single resource service Agent is the result of present application collaborative manufacturing resource packaging. Considering reducing man-made modified work, when application resources are realizing application service, Agent has to be packaged easily and keep simple structure. The message router server accepts browsing requests, and sends requests to the corresponding application service Agent to deal with. The unit knowledge recorded in MAS give the message router server reasoning ability, to sort, for example, user requests, and to combine feedback.

The Logic Structure of Collaborative Manufacturing Resource Sharing Original System in Network Environment

Front and back sub-systems are both based on Web technology, taking a mixed structure of B/S and C/S. The collaborative manufacturing resource is based on a relational database management system, collaborative manufacturing resource categories and collaborative manufacturing resource expressing model constituting collaborative manufacturing resource library, and it is on the lowest layer of the original system. The logic structure of the original system in the network environment is shown in Figure 7. It takes the Internet Web server as its center, to develop different application service programs for each unified enterprise based on the Web server, and those application service programs will be distributed to application servers after being distribution object packaged, and visiting the database server through ODBC or JDBC. United enterprises, with the support of the server's application tool sets, link with the server through the browser to use various kinds of information communication and share functions that the resource sharing service platform

Figure 7. The logic structure of collaborative manufacturing resource share original system in network environment.2 function model design of system



provides. Function models of Collaborative manufacturing resource share website include Collaborative manufacturing resource service, website navigation, dynamic collaborative union, message board, Collaborative manufacturing resource request information; function models of Collaborative manufacturing resource maintain management system include dynamic information management, safety management, website management, system management and customer information management.

The Design of the System Modular Function

The following content explains the resources and network service functions that a collaborative manufacturing resource sharing system needs to integrate from a logic structure view, which contains ten aspects:

1. **Collaborative manufacturing resource service:** mainly provides modeling,

integration, package, search and restructures for collaborative manufacturing resources. Collaborative manufacturing resource modeling provides functions such as resource property input resource category and resource model modification. Collaborative manufacturing resource integration provides model match definition modification and deletion of mapping relation. Collaborative manufacturing resource package provides CAD, ERP, PDM and other types of resource package functions. Collaborative manufacturing resource search provides many searching tools such as enterprise key words searching, resource category searching and roughness match searching, so that enterprises can localize the resources they need. Collaborative manufacturing resource structure service can choose new resources according to enterprise requirements when an enterprise's resources can meet its needs.

2. **Collaborative manufacturing resources requirement information:** to publish on the Internet, to ensure that enterprises in different places can understand the latest information of collaborative manufacturing resources, to maximize profit, including a requirement publish and requirement overview. Display requirement information according to resource categories, to make a single list for highest-clicked requirement, and to collect clicking times for each requirement information.
3. **Website navigation:** this is mainly to help customers understand system functions as quickly as possible and to help them use the system efficiently. It includes system introduction system functions system characteristic user direction and FAQ.
4. **Dynamic collaborative alliance:** these are membership enterprises that publish project unified information to global potential partners according to their own manufacturing ability. It includes alliance building function, password function, alliance management function, expert library management function and expert evaluation function.
5. **Message board:** this is mainly for the online communication of collaborative manufacturing resource information, as with BBS. It includes reading messages, publishing messages and deleting messages.
6. **Dynamic information management:** this is a system administrator that adds or deletes news and requirement information, checks and deletes messages on message board.
7. **Safety management:** this is a function of the collaborative manufacturing resource sharing system that is vital. It contains the system visit control, firewall configuration and anti-virus configuration, copy of system data and membership enterprise login and the ID check of logging out.
8. **Website management:** this takes care of the management of the collaborative manufacturing resource sharing system. The network diary contains various diary documents of related Web visits that are stored on the server; website page views and advantage resources are analyzed by exploiting the Web diary, so as to improve users' visiting effectiveness. Collaborative websites introduce some similar manufacturing resource sharing platforms or related websites, which assists collaborative enterprises and users to learn various information about product manufacturing. User management is the management of user information.
9. **System management:** its main function is to ensure that the system administrator delivers different visiting right degrees for different level users; to maintain daily the stable running of the collaborative manufacturing resource share system; to update various data in the database. In addition, it has the function of adding, modifying and

deleting the system administrator, that is, the system administrator management.

10. **Customers information management:** this provides membership enterprise registration, the deletion or modification of login password, information modification, to collect fee online and get feedback.

Conclusion

In the case study, the method and framework of the collaborative resource share based on multi-agent are proposed. Collaborative manufacturing resources scattered over the country are encapsulated as agents, which join the collaborative manufacturing resource chain in the form of node. So the whole world of manufacturers is viewed as a virtual network manufacturing system, which can realize manufacturing resources share and cooperative manufacture.

SUMMARY

In this chapter, an overview and basic theory of Intelligent Agents and MAS were introduced, including their definitions, properties, and architectures. The chapter presented the development history of agent and MAS, summarized the current research status of MAS and Intelligent Agent, and discussed practical Agent technologies, including communications and interaction, and collaboration and behavior coordination between agents. The international organization FIPA and the development tools of MAS were also introduced. Finally, a case study about logic architecture of collaborative manufacturing resource share prototype system in net environment was studied.

Recent development in agent technology produces new and interesting opportunities. Agent technology plays a significant role in the development and implementation of complex systems and is in the mainstream of software development technologies. A large number of successful applications of Intelligent Agents have appeared and many new application areas are still emerging. There are still many problems to address as follows including the Agent's own inherent defects:

- Systems developed by Agent technology will be uncertain. Agents are automatic and can make decisions freely to some degree, so the relationship between Agents cannot be accurately forecasted and managed.
- Properties and behaviors of the system cannot be determined during the design period of the system. Though the designer can establish the code of conduct for individual Agents, the overall behavior only is realized during operation.
- The correspondent software engineering for MAS is still immature, which has largely hindered the in-depth development of Agent technology. At present, the relevant theoretical system of Agent technology is incomplete, and it is the trend that future agent technology will pay great attention to enriching and improving the theoretical system and development methods of agent technology. Agent technology can currently only be utilized by high quality technical engineers, so how to promote the relevant development methods and technologies is an important issue requiring attention when Agent technology is utilized.

REFERENCES

Boonserm, K., Richard, A. W., Hyunbo, C., & Albert, J. (2004). Integration Framework of Process Planning based on Resource Independent Operation Summary to Support Collaborative Manufacturing. *Int. J. Comput. Integ.Mfg*, *17*, 377–393. doi:10.1080/09511920310001607104

Chapman, D. (1987). Planning for Conjunctive Goals. *Artificial Intelligence*, *32*(3), 333–377. doi:10.1016/0004-3702(87)90092-0

Ferber, J. (1998). *Multi-Agent Systems: Towards a Collective Intelligence*. Reading, MA: Addison-Wesley.

Ferber, J. (1999). *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Reading, MA: Addison Wesley Longman.

Fikes, R., & Nilsson, N. (1971). STRIPS: A new approach to the application of theorem proving to problem Solving. *Artificial Intelligence*, *2*(3), 189–203. doi:10.1016/0004-3702(71)90010-5

FIPA. (n.d.). *Communicative Act Library Specification*. Retrieved from http://www. fipa.org/specs

FIPA. (n.d.). *Interaction Protocol Library Specification*. Retrieved from http://www.fipa.org/specs

Foundation for Intelligent Physical Agents. (n.d.). *FIPA 97 Specification. FIPA Architecture Overview*. Retrieved from http://www.fipa.org/specs

Franklin, S., & Graesser, A. (1996) Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages.* Berlin: Springer-Verlag.

Gaines, B. R., Norrie, D. H., & Lapsley, A. Z. (1995). An Intelligent Information System Supporting the Virtual Manufacturing Enterprise. In *Proceedings of 1995 IEEE International Conference on Systems, Men and Cybernetics. Vancouver*, 10.

Gao, B., Fei, Q., & Chen, X. G. (2001). Agent Interactive Model. [In Chinese]. *Computer Science*, *28*(8), 105–109.

Genesereth, M. R., & Ketchet, S. P. (1993). Software Agents. *Communications of the ACM*, *37*(7), 48–53. doi:10.1145/176789.176794

Hayes-Roth, B. (1995). An Architecture for Adaptive Intelligent Systems. *Artificial Intelligence: Special Issue on Agents and Interactivity,* 329-365.

Hewitt, C. (1973). A Universal Modular Actor Formalism for AI. In *Proceedings of the 3rd International Joint Conference on Artificial Intelligence (IJCAI-73)*, (pp, 235-245).

Jennings, N. R. (2000). On Agent-based Software Engineering. *Artificial Intelligence*, *117*(2), 277–296. doi:10.1016/S0004-3702(99)00107-1

Jennings, N. R., & Wooldridge, M. J. (1995). Intelligent Agents: theory and practice. *The Knowledge Engineering Review*, *10*(2), 115–152. doi:10.1017/S0269888900008122

Jennings, N. R., & Wooldridge, M. J. (1998). Applications of Intelligent Agents . In Jennings, N. R., & Wooldridge, M. J. (Eds.), *Agent Technology: Foundations, Applications, and Markets* (pp. 3–28). Berlin: Springer.

Jiao, J. R., You, X., & Kumar, A. (2006). An Agent-based Framework for Collaborative Negotiation in the Global Manufacturing Supply Chain Network. *Robot. Comput. Integr. Mfg*, *22*, 239–255. doi:10.1016/j.rcim.2005.04.003

Lacey, H., & Deloach, S. A. (2004). Automatic verification of Multi-Agent conversations. In: 8th Annual Midwest Artificial Intelligence and Cognitive Science Conference, University of Arkansas, Fayetteville, AAAI Press, (4): 93~100.

Liu, S., & Young, R. I. M. (2004). Utilizing Information and Knowledge Models to Support Global Manufacturing Co-ordination Decisions. *Int. J. Comput. Integ. Mfg*, *17*, 479–492. doi:10.1080/09511920310001650025

Maes, P. (Ed.). (1990). *Designing Autonomous Agents: Theory and Practice From Biology to Engineering and Back*. Cambridge, MA: MIT Press.

Minsk, M. (1986). *The Society of Mind*. New York: Simon & Schuster.

Minsky, M., & Riecken, D. (1994). A Conversation with Marvin Minsky about Agents. *Communications of the ACM*, *37*(7), 23–29. doi:10.1145/176789.176791

Nahm, Y. E., & Ishikawa, H. (2005). A Hybrid Multi-Agent System Architecture for Enterprise Integration Using Computer Networks. Int. J. Robot. Comput.- . *Integr. Mfg*, *21*, 217–234.

Nilsson, N. J. (1998). Artificial Intelligence: A New Synthesis. San Francisco: Morgan Kaufmann Bradshaw, J.(Eds) (1997). Software Agents. Cambridge, MA: AAAI Press/MIT Press.

Nwana, H. S. (1996). Software Agents: an overview. *The Knowledge Engineering Review*, *11*(3), 205–244. doi:10.1017/S026988890000789X

Qiao, B., & Zhu, J. Y. (2001). Multi-Agent Manufacturing Systems. Nanjing University Journal of Aerospace, 33(1): 1-7. (In Chinese) Rosenschein, J. S. & Zlotkin, G. Rules of Encounter: Designing Conventions for Automated Negotiation among Computers. MIT Press, Cambridge, MA.

Shen, W., Hao, Q., & Norrie, D. H. (2006). Applications of Agent-based Systems in Intelligent Manufacturing: An Updated Review . *Advanced Engineering Informatics*, *20*, 415–431. doi:10.1016/j.aei.2006.05.004

Shen, W., & Norrie, D. H. (1999). Agent-Based Systems for Intelligent Manufacturing: A State-of-the-Art Survey. *Knowledge and Information Systems*, *1*(2), 129–156.

Shen, W., & Norrie, D. H. (2001). *Multi-Agent Systems for Concurrent Intelligent Design and Manufacturing*. London, UK: Taylor and Francis.

Shen, W., & Norrie, D. H. (2001). Dynamic Manufacturing Scheduling Using Both Functional and Resource Related Agents . *ICAE*, *8*(1), 17–30.

Shi, Z. Z. (2000). *Intelligent Agents and its Applications*. Beijing: Science Press. (In Chinese)

Swaminathan, J. M., Smith, S. F., & Sadeh, N. M. (1996). A Multi-Agent Framework for Supply Chain Dynamics, in: Proc. of AI & Manufacturing Research Planning Workshop, Albuquerque, NM.

Wang, D., & Nagalingam, S.V. 7 Lin, G.C.I. (2005). (in press). Development of an Agent-based Virtual CIM Architecture for Small to Medium Manufacturers. Robot. Comput.- . *Integ. Mfg*.

Wooldridge, M. J., & Jennings, N. R. (1998). Pitfalls of Agent-Oriented Development. Proceedings of the 2nd International Conference on Autonomous Agents. New York: ACM Press. pp: 386-391.

Chapter 4
Data Mining and Knowledge Discovery

In Chapters 2 and 3, the knowledge-based system and Multi-Agent system were illustrated. These are significant methods and theories of Manufacturing Intelligence (MI). Data Mining (DM) and Knowledge Discovery (KD) are at the foundation of MI. Humans are immersed in data, but are thirsty for knowledge. With the wider application of database technology, a dilemma has arisen whereby people are 'rich in data, poor in knowledge'. The explosion of knowledge and information has brought great benefit to mankind, but has also carried with it certain drawbacks, since it has resulted in knowledge and information 'pollution. Facing a vast but polluted ocean of data, a technical means to discard the bad and retain the good was sought. Data Mining and Knowledge Discovery (DMKD) was therefore proposed against the background of rapidly expanding data and databases. It is also the result of the development and fusion of database technology, Artificial Intelligence (AI), statistical techniques and visualization technology (Fayyad U., 1998). DMKD has become a research focus and cutting-edge technology in the field of computer information processing (Jef Woksem, 2001). The development background, conception, working process, classification and general application of DM and KD are firstly introduced in this chapter. Secondly, basic functions and assignment such as prediction, description, data clustering, data classification, conception description and visualization processing are discussed. Then the methods and tools for DM are presented, such as the association rule, decision tree, genetic algorithm, rough set and support vector machine. Finally, the application of DMKD in intelligent manufacturing is summarized.

DOI: 10.4018/978-1-60566-864-2.ch004

INTRODUCTION

Background and Conception

DMKD originated from Knowledge Discovery in Database (KDD). It first appeared in August 1989 at a meeting of the 11th International Joint Conference of Artificial Intelligence. KDD has been defined as the non-trivial process of identifying hidden, previously unknown and potentially useful information

from data (Frawley W., Piatesky-Shapiro G. & Matheus C.,1991).

There are many similar ways of describing KDD, such as Data Mining (DM), Knowledge Extraction, Information Discovery, Information Harvesting, and Data Archaeology. Although definitions differ, the essence of extracting hidden, interesting and high-level models from the data is uniform. Of these names, KDD and DM are two commonly-used terms. Scholars in the field of statistics, data analysis and information systems regular employ 'DM', but experts in the field of AI and machine learning use 'KDD'. DM, etc.

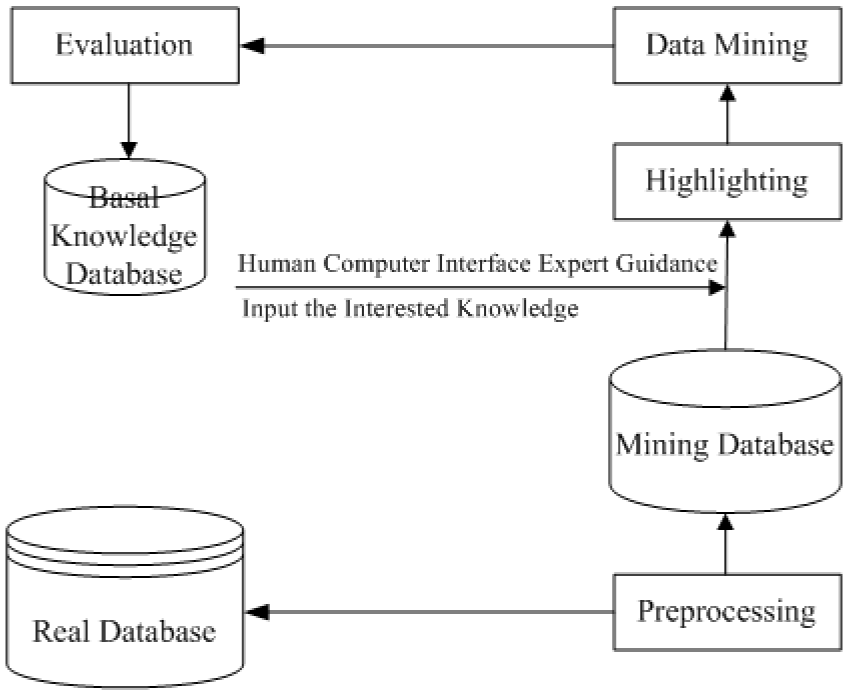
In order to unify understanding, Fayyad offered a new definition of KDD and DM and made a distinction between them in a field paper (Fayyad U. & Piatesky-Shapiro G. et al., 1996). The new definition for KDD is this: KDD is a process of identifying effective, innovative, potentially useful and ultimately understandable models. DM is defined as: DM is a step that generates a specific mode through a specific algorithm in an acceptable computational efficiency. KDD is an entire process including the steps of data choice, data preprocessing, data transformation, DM, and pattern evaluation, which eventually lead to obtaining knowledge. DM is one of the key steps. Despite the distinction, scholars often equate KDD with DM for convenience. When referring to DM' alone, it is taken to also mean the data preprocessing and the results of evaluation. Consequently, people tend to use these two names together, so DMKD occurs.

In simple terms, DM is extracting or mining knowledge, and can be defined from the point of statistics, databases, machine learning and so on. 'Mining' first appeared in statistics, from which perspective, DM is analyzing the data sets observed in order to find the unknown relationship between data, and providing understandable, innovative and useful summarized data to the data owner (Han J.W. & Kamber M., 2002). From the perspective of databases, DM is a process of finding interesting knowledge from large amounts of data stored in databases, data warehouse and other information storage (Han J.W. & Kamber M., 2001). From the perspective of machine learning, DM is defined as extracting the connotative, obviously unknown and potentially useful information from data (Witten I. H. & Frank E., 2000).

From 1989 to date, the DMKD international conferences sponsored by the American Association for Artificial Intelligence have been held many times. The conference was a biennial symposium before 1993, but since 1995 it has been held once a year and has developed into the international academic conference named KDD. The attendance increased from dozens to hundreds, and the ratio of accepted papers rose from 2:1 to 6:1. In 1997, the first Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDDD) was held in the Asia-Pacific region and has since been held annually. The DMKD academic conferences have also been held once a year in Europe since 1997 under the name of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD). In recent years, relevant international conferences such as Very Large Data Bases (VLDB) and Special Interest Group on Management of Data (SIGMOD) have also attracted a large number of DMKD papers, reports and corresponding topics. A DMKD special has been published in the IEEE Journal of Knowledge and Data Engineering, Intelligent Systems, Computational Intelligence and other journals in recent years. The international DMKD journal 'Data Mining and Knowledge Discovery' was first published in January 1997.

So far, much progress has been achieved in the research of DM and KD in relational databases and service databases. DMKD technology was initially application-oriented, because it originated against a background of strong application demand. In the international community, DMKD technology has been used in marketing, finance and banking, insurance, telecommunications, transportation, and other fields. In recent years, the spatial database,

Figure 1. Process of KDD



temporal database, multimedia database, and web data excavation has attracted wide attention. The depth of research on DMKD theories and methods has been continuously strengthened, and at the same time the extent of application has been expanded. DMKD has been one of the most popular topics in the field of information technology in the 1990s and into the 21st century.

In general, DMKD is equivalent to the position of database technology in the 1970s. The guidance of theory and methods is desirable, and the models and tools which are similar to the Database Management System (DBMS) system and the SQL query language are necessary too. The application of DMKD is promoted universally according to these tools. Spacial DM and KD are now started at the initial stage, and there is an urgency to create systems theory and technical frameworks, and to develop effective algorithms to start a typical application.

Process of KDD

KDD is a multi-step process which includes the following major steps. The process of KDD is illustrated in Figure 1. These steps may be repeated many times in the process.

1. **Data Selection:** According to the user's requirements, data relevant to KDD is extracted from the database. KDD will extract knowledge from this data. In this process, it makes use of the database knowledge to deal with data, and establish database.
2. **Data Preprocessing:** Data produced in step (1) is preprocessed and checked for integrity and consistency. The noise data is processed and the loss data is filled using statistical methods to form a mining database.
3. **Goal of KDD:** Because it uses different algorithms to the different requirements of KDD in specific KD process, the type of

knowledge that found by KDD is determined according to the user's requirements.

4. **Determining the KD Algorithm:** According to the task determined by step (3), the appropriate KD algorithm is selected, including a suitable model and parameters. This makes KD algorithm and the standard of judgments of KDD difficult to be consistent.
5. **Highlighting:** Highlighting means data is selected from the mining database. The method of highlighting mainly uses cluster analysis and discriminating analysis. Experts input the interested knowledge through the human-computer interaction and guide the direction of DM.
6. **DM:** Using the selected KD algorithm to extract the required knowledge, these could be shown in a specific manner, or by a common approach, such as a generating rule.
7. **Evaluation of knowledge:** The process used to evaluate the value of the rules obtained to determine whether the rules are inputted to basic database. Conducted largely through human-machine interfaces, experts rely on experience to evaluate.

From the above introduction, it can be seen that DM is only one step in KDD which mainly uses certain KD algorithms to discover the relevant knowledge from the data within certain limits of operational efficiency.

The several steps of the whole process of KDD are further summarized into three parts. They are DM pretreatment (the preparatory work before DM), data mining (DM) and DM post-processing (the processing work after DM).

Classification of DM

DM is a field which is cross-disciplinary in nature, involving database technology, statistical theory, machine learning, pattern recognition technology, and the theory and technology of visualization. Different DM systems are proposed according to different methods of DM, different mining data and knowledge, and different DM applications of. Mastering the different classifications of a DM system is helpful to people to determine the most appropriate DM system. The classification of DM is discussed as follows.

1. Classification of data models includes the relationship-based DM system, object-based DM system, object-relationship DM system, affair-based DM system, data warehouse-base DM system, and so on.
2. According to the classification of the specific type of data handled, there are deduction DM systems, space DM system, time-sequence DM systems, space-time DM system, multimedia DM system, text DM system, WWW(World Wide Web) DM system, and so on.
3. The different types of knowledge mined could be divided into characterization, discrimination, association, classification, clustering, isolated point analysis (abnormal data), evolution analysis, deviation analysis, and similarity analysis.
    a. A characterization formula is extracted from a group of data related to learning tasks in the characterization rules DM systems. A characterization formula includes the overall characters of the data sets. The overall characters collect the characteristics rules set that are hidden in the objective database.
    b. Certain attributes or characteristics of learning data (objective data) are discovered and extracted to distinguish the learning data from comparative data in a discrimination rules DM system. The chanciness of data that implied in the database and the trend related to a specific model are collected to build the similar matching rules of distinguished model.

c. Relation and association among a group of projects through relevance are discovered in an association rules DM system, and these relationships are expressed as rules; that is, sets of association rules are collected from the affair database and relationship database.
d. Large classifications of data and collecting corresponding sets of classification rules are produced in a classification rules DM system.
e. A limited type of collection or cluster set is searched and identified in a clustering rules DM system to describe data. Clustering also means clustering a data set based on the concept of the clustering principle (identifying a group of clustering rules) to polymerize similar affairs together.
f. An isolated point is a data object that is out of line with the general data pattern. Isolated point analysis is mining these points.
g. Significant changes and deviations among the status quo, history and standard are detected in a deviation analysis rules DM system. Thresholds from different concepts of test are collected to form the set of detection rules; a comprehensive DM system should provide a wide range and integrated function of DM.

4. According to the granularity and the level of abstraction of the excavated knowledge, a DM system is a classified general knowledge mining system, which collects the generalization knowledge (the high level of abstraction) of the data that is hidden in the objective data set, the original level knowledge mining system, which collects the regularity of data which is hidden in the layer of the original data (layer of the original data), and the multi-level knowledge mining system, which is collects the knowledge in a number of abstraction layers. The DM system is also classified as mining the regularity of data (general model) and irregularity of data (such as unusual or isolated)). An advanced DM system supports the more abstract level KD.
5. According to the technology used in DM, there is the automatic DM system, confirmed driver DM system, found drive DM system and interactive DM system.
   a. **Automatic DM system:** Unknown and useful models are discovered automatically from the large amounts of data in this DM system. It is the advanced level of DM.
   b. **Confirmed driver mining system:** Users create assumptions (or models) based on experience, then test the assumptions (or mine the data that matches the models) through a confirmed driver operation. The testing process is the DM process. The information that has been extracted may be the trend or the fact. The operations of confirmed driver DM are query and reporting, multi-dimensional analysis and statistical analysis. Among them, the purpose of query is to express a hypothetical effectively. The report is the description for the results of analysis. Multi-dimensional analysis uses specific query language and visualization tools to analyze the hierarchical structure of each dimension. Statistical analysis construes the data by combining the statistical with DM and visualization technology.
   c. **Discovery driver mining system:** Creating a model to predict the future on the target data set by using historical data. The process of creating a model is the process of DM. The knowledge that is mined is likely to be

a regression or classified model, due to the relationship between the database records, error situation, and so on. The operations of the discovery driver mining system are the forecasting model, database segmentation, link analysis (associated discovery) and deviation detection. Forecasting modeling is the basic operation of the discovery driver mining system. According to past behavior, the operation generates a model automatically by using the historical data to predict future behavior. The operation of database segmentation classifies the database as a related collection of records, gathering classified data. The technology that supports database segmentation is clustering. If modeling and separating operations are general descriptions of the establishment of data, then link analysis is the relationship between records of the establishment of the database. In other words, some relationship often is hidden between the records in the database, reflecting certain relevance. Link analysis identifies the relevant DM operation. The technology that supports link analysis is joint discovery and association discovery. The purpose of deviation detection is to identify the data that is not suited to division, then note that this data is the noise, or the data that needs to do detailed testing, and the operation is related to database segmentation. The main technology to support deviation detection is statistical technique.

6. According to the different methods of data analysis (such as database-oriented approach, data warehouse-oriented approach, machine learning approach, statistical methods, pattern recognition, or neural network), there are different classifications. The mining system based on summary is a system of summarizing the general characteristics and high-level knowledge of the designated objective data by using the tools for concluding and summarizing data. The model-based mining system is a system of mining the data that matches with the model according to the prediction model. The mining system that is based on statistics is a system that mines objective data according to statistical theory. The mining system that is based on the mathematical theory, as the name suggests, mines the objective data based on mathematical theory. The integrated mining system consolidates a variety of DM methods on objective DM.
7. The DM system is classified according to its application. The corresponding DM industries are as follows: financial, telecommunications and insurance industries, commercial, manufacturing, sports, astronomy, medical and health fields, and the transport field.

General Application of DM and KD

DM has taken only a few years to develop from the theoretical research stage to product development, and it has now entered the application stage. The application of DM technology is very broad. It is applied in the area of government management decision-making, business, scientific research, industry decision-making support and many other fields.

Customer Database-Based Marketing

Customer database-based marketing is the first and most important area of DM technology. It is the first step in marketing to understand the customer's situation. DM technology is regarded as one of the main supporting technologies in customer relationship management (CRM). In the true sense of the 'one-on-one' marketing tool, the purpose of CRM

changed from the traditional role of obtaining new customers to retaining existing customers, and from gaining market share to obtaining customer share. The targets of CRM are retaining profitable customers through better services, winning new customers through lower costs, expanding market share, giving up non-profitable and poor credit customers, reducing operating costs and risks. Specific factors include the overview of customer analysis, customer loyalty analysis, the composition of groups of customers, customer consumption levels, the largest contribution to the client, good or bad credit customers, geographical distribution of customers, customer's spending habits and potential consumer demand.

Customer databases are the central repository that contains all the information about a customer. It is the analysis of the DM sources of information, records of customers and business contact information, including customer purchase transactions, telephone enquiries, evaluation, returns, telephone service complaints and general complaints,. In addition, it includes personal information about customers such as age, income, marital status, children, place of residence, interests, and so on. Customer data comes from credit cards, membership cards, website registration forms, as well as from customer complaints about products and services, and various other sources. In the field of marketing, the terminal use of DM is locating targets and customer descriptions. By locating targets, it is possible to find customers with similar buying patterns and characteristics, and to take action promoting specific commodities to them. It is also possible to identify the characteristics of customers of rival companies, which is helpful to companies in adopting measures to retain customers who might otherwise be lost. The characteristics of these types of customer are found and predicted in customer descriptions via DM techniques.

Retail Marketing

DM technology helps businesses to identify the habits of consumers, predict customer buying patterns, partition groups of customers, find the link between sales of goods, optimize placement, research the best time for promotional activities and the promotional merchandise mix, and to understand poor sales and sales status through analyzing the customer shopping basket. It is helpful to enterprises to establish an effective strategy of sales and advertising through the market share analysis of a commodity in the chain, customer statistics and historical analysis of the situation.

Credit Card Marketing

Credit card use is a consumption pattern that is widely seen in business. In order to meet increasingly fierce competition in the market, credit card companies often introduce incentives such as frequent flyer mileage. Such concessions are not indiscriminate, but developed for certain specific customers. Because it involves customer privacy, airlines can not directly purchase the name, address and other information relating to the customer from credit card companies, but useful information can be obtained, such as the names of customers who spent more than 2000 dollars on flights in the last six months. For these customers, credit card companies and airlines combine to provide free travel and other concessions.

Telecommunications Marketing

In telecommunications companies, DM techniques are adopted to help enterprises to analyze the user's experience of using the telecommunications service, to analyze customer behavior, and to optimize advertising investment. Large amounts of data are accumulated in the process of telecommunications operations, including basic customer information, customer contact decision-making, customer interaction records, sales costs, marketing state, products, prices, trends, call-center interactive

records and account information. This information is very important in enhancing insights into customer behavior, exploring customers' needs, improving customer satisfaction and loyalty, and prospecting for new customers by using customer trends and patterns of consumption. CRM in telecommunications companies mainly includes sales process management, sales funnel management, order management and sales analysis Sales process management is managing the interaction data of business staff and customers. Sales funnel management is managing different states of customer, expected list quantity, sales with list and so on. Order management is managing the products and services purchased by customers, the price paid, and the period in which the products or services were accessed. Sales analysis is analyzing the sales situation by product, staff time, prices and trends.

Others

In the insurance industry, DM methods are used to analyze the main factors of insurance, the risk model, and to launch different types of insurance for customers. In the production and marketing of enterprises, DM predicts sales and inventory demand, and price analysis. For example, the goods are classified into several levels to help shopping malls for with market and pricing of goods: 'popular and high profit', 'best-selling and low profit', 'best-selling non-profit', 'not popular and high profit' and 'non-profit and not-selling'.

Risk Analysis and Fraud Detection

DM is also useful in risk assessment, financial planning and asset evaluation, resource planning, competitive strategy choice and so on.

Insurance Risk Analysis

Because of the great risks in financial investment, the best direction in investment decision-making is chosen by analyzing the various investment directions of related data.

Risk estimation plays an important role in guiding the normal work of insurance companies, the premiums charged and the design of the guarantee slip, all of which need detailed risk analysis. Generally, premiums are determined after analyzing major factors such as the policyholder's age, gender, health, number of children and so forth. Useful rules are also obtained through using DM to analyze risk investment in determining the appropriate levels of compensation.

Financial Industry Risk Analysis

Financial services need to collect and process large amounts of data, analyze this data, establish data models and features, find the financial and commercial interests of customers, consumer groups or organizations, and observe the changes in financial market trends. In the financial sphere, the bank's special status, nature of work, business characteristics and fierce market competition has exerted a more pressing demand for information technology and electronic support than other areas. Using DM technology can help banks with product development by describing the demand trends of the past and predicting the future. DM technology has been used widely in several commercial banks in the United States.

Fraud Detection

Recognition of credit card fraud has been looked upon as the main application of DM technology for credit card companies. Different usage models are found by analyzing credit card data. Over a relatively long period of time, a customer's credit card use habits are often fixed. Exceptional or unusual credit card usage can therefore be detected and legal users are identified according to these credit card usage models.

The Application in the Field of Sports

Advanced Scout is a data analysis tool used, for example, by the National Basketball Association (NBA). Coaches utilize a portable computer to

mine the data which is stored in the NBA's central server. All activities in every match are classified statistically by scores, secondary attack and errors. Coaches easily understand the rules which are found by Scout through searching the previous NBA games video by using time mark. There are already some 20 NBA teams using IBM's development of the DM application software-Advanced Scout system to refine their tactics portfolio and to make decisions about replacing team members.

BASIC ANALYSIS

Basic Function-Prediction and Description

There are two general functions of DM-prediction/verification and description. Unknown database characteristic values are predicted and verified via several known attributes of data in the former function. In the latter function, the understandable patterns of data description are discovered. Prediction also encompasses the identification of distribution trends based on the available data. Classification and prediction may need to be preceded by relevance analysis, which attempts to identify attributes that do not contribute to the classification or prediction process. These attributes can then be excluded.

Function of Prediction/Verification

Prediction is forecasting unknown numerical or loss of data through analyzing and processing known data - for example, predicting a worker's wages according to his colleagues' wages. Trend analysis is the attempt to predict patterns or trends based on past data. According to previous behaviors, the possible future behavior can be predicted. Common temporal data include supermarket transaction records, hospital medical records and weather data. Changing patterns of behavior can be obtained through trend analysis, such as predicting the trend of stock movement according to the changes in stock levels in a given period.

There are many trend analysis and forecasting methods, some of which are discussed below.

Statistical Methods

Statistical methods are useful for identifying a property relative to the predicted value. The distribution of property values is estimated according to the analysis of similar data. Regression analysis is a common method of prediction. This describes the relationship where properties Y depend on the properties set of X. Y is known as the target property, X is called the conditions property. The use of regression analysis methods can show the dependent relationship between the target property and conditions property; not only the relationship between the variables is described, but the prediction is also accomplished based on this relationship.

Method Of Association Rules

Target property values are transformed into discrete values through domain compartmentalizing. The method of association rules are used to predict the property values of numerical goals.

Association rules describe the attributes of objective property through the testing of condition properties and predict the target property value by those rules. Assuming Y is the target property, X is the conditions property testing, the association rules $X \Rightarrow Y = a$ shows that if the group of t meets the testing X, the possibility of t equivalent to a is higher.

Decision Tree and Regression Tree Method

A decision tree is a tree with a root. Each internal node contains the testing of a conditions property. The element groups arrive at leaf nodes through the cycle testing of internal nodes. The value of leaf nodes is regarded as the predicted value of the target property of element groups.

A regression tree is a binary tree with a root. Each internal node contains the testing of a conditions property. Unlike the decision tree, the target property of the regression tree is numerical. The average value of a property of a leaf node is regarded as the predicted value of the numerical target property.

Average variance, which is the true value of the corresponding property of each element group and the average value of the square deviation of the predicted value, is usually used to estimate the accuracy of the regression tree. The structural idea of the regression tree is to split the corresponding node into two sub-categories every time according to some selective testing, then to add it respectively to the left and right sub-nodes. After adding the new node into the return tree, the average variance of the regression tree becomes smaller. Eventually a regression tree with a minimum average variance is generated.

Sequence Pattern Mining Method

The existence of temporal data makes it necessary to considering the time factor in the process of KD. A sequential pattern is the expansion of association rules on the sequence of time events. Its purpose is founding the continual occurring issue sequence from amount issues sequences. Sequence model mining can be used to assess the relevance of the sequence of events in the socio-economic context, and to predict future sales models. It can be seen that events sequence is a common data format in daily life. If every customer's single purchase in shopping malls is regarded as an event, then the number of purchases of this customer over a period of time forms a purchase events sequence. The important issue of KD in the sequence of events is in discovering the events sequences which often appear from multiple events sequences.

Function of Description

Data Classification

Data classification is used to find the same property of a group of objects in the database and divide the property into different categories according to their classification model. Data classification is a process that identifies the map functions between each type of data and establishes categories. It is widely used in such fields as medical diagnosis, market research and credit assessment.

Assuming that the database record is described by a group of characteristics, each record is given a type mark. Such data is called a sample database or training set. Classification is creates an accurate description for each category by analyzing the data in the sample database, establishing a classification model, then classifying other unknown types of records by using this classification model.

Classification is a supervision learning method for constructing the classification model. It usually chooses a number of example data from the database as training sample data sets. Others are taken as a testing sample data set. According to the training set analysis, a model is built that shows their characteristic. A precise description of each category is then obtained. Finally it is extended to a better classification model through the data of the testing set. Common classification methods are decision trees, neural networks, genetic algorithms, support vector machine, rough set theory, K-neighbor algorithm, Bayes classification, and simple distance classification.

Classification is the process of finding a set of models (or functions) that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data (i.e., data objects whose class labe1 is known).

Classification can be used for predicting the class label of data objects. However, in many applications, users may wish to predict some missing or unavailable data values rather than

class labels. This is usually the case when the predicted values are numerical data and is often specifically referred to as prediction. Although prediction may refer to both data value prediction and class label prediction, it is usually confined to data value prediction and thus is distinct from classification.

Clustering Analysis

Clustering analysis is the process of dividing a group of objects into a series of meaningful subsets according to property values. Input of clustering is a group of unmarked data. Data are classified according to their own distance or similarity. The principle of dividing is to maintain the largest similarity in the group and the smallest similarity between the groups. This method breaks down large-scale systems into smaller components, thus simplifying the system design and implementation.

Unlike data classification, the input of clustering classification is a group of unclassified data. Sometimes there is no clear idea about the classification of these data in advance. Clustering divides the data set reasonably through the analysis of these data and the similarity between them. There are many ways of clustering analysis, including level clustering, divide clustering, density clustering, fuzzy clustering and grid methods. For the same data set, different results are obtained by using different methods of clustering. Therefore, evaluation of clustering results is an important step of clustering analysis.

Data Collection And Induction

The conceptual level is the general existing structure in database. It is used to describe the affiliation among data. Data in the database often contains detailed information about the original concept. The process of summing up a data set into the concept of high-level information is called data collection. There are two main ways to do this.

The data cube method calculates regularly queried and costly operations first and stores them in array.

The attribute-oriented induction method expresses the query requirements by DM query language which is similar to SQL. Data are collected according to attribute deleting, conceptual level tree, threshold controlling, amount communication, collection function and so on. Eventually they are transformed into rules, such as characteristics rules, discriminate rules, classification rules and related rules.

Summarization

The method of summarization is finding the description of data collection, available charts, association rules and visualization technology.

Dependency Pattern

The method of the dependency pattern is constructing the function-dependent relationship or correlation model between variables.

Deviation Detection

Deviation detection is finding the deviation between the results and expectations through testing differences between data status and history records or standards.

Data Clustering and Classification

In everyday human life, it can be observed that a baby can easily distinguish its mother among many people. This is an example of a classification learning process. When the number of objects being studied is particular large, this process often is called a study based on DM.

According to whether there is supervision in learning, learning methods based on DM are divided into two categories. They are supervision learning and no supervision learning. The former means that the knowledge is known. The construction of the model is guidance, and the purpose of learning is to construct an optimal

model and minimize the errors between the known knowledge and the model. It usually takes the form of the optimizing process. The latter – no supervision learning - is different. The knowledge to be found is unknown. The model is entirely extracted from data and the accuracy of the model is difficult to verify.

Data classification (simplified classification) is a typical supervision machine learning method. Its purpose is to find the classification model from a group of data of known types and predict the unknown category of the new data. Prediction is another supervision learning style. Generally, the difference between prediction and classification is that predicted things are values rather than categories.

Data classification technology is widely applied in credit card approval, target market positioning, medical diagnosis, exception detection effectiveness analysis, graphics processing and insurance fraud analysis, among other fields. Classification is used extensively in commercial, banking, insurance, medical sanitation and scientific research

The application in medical diagnosis is first discussed. When a doctor diagnoses the patient's condition, he first looks at the patient's symptoms and determines the type of disease, then proposes the right remedy. Although much advanced equipment and technology has become available with the development of modern medicine, and it is easy to get all a patient's details, still medical diagnosis is a complex issue that requires comprehensive analysis of almost all the symptoms of the disease to ultimately determine the type. In particular, if the patient's symptoms are not clear or different types of disease pathology present similarly, misdiagnosis may occur even by experienced doctors. This results in the patient missing the optimal time for treatment. Classification is used for medical diagnosis, and the key features of some diseases from a large number of clinical cases are used to help doctors make more accurate diagnoses and improve valuable treatment time for patients. Apart from use in expert fields, classification is also used in daily life. For example, weather conditions are classified into the 'suitable' and 'not suitable' types on the basis of past wind, humidity, temperature and sunshine, and other weather, and are used to guide people's everyday activities. Classification technology is used by astronomers to find rare celestial bodies or stellar spectra, to determine the form of classification of APM galaxies, or to make the distinction between stars and galaxies and galaxies of different activities, such as nuclear or non-nuclear.

Data used for classification are a group of known type samples (also called data, objects, cases or records). Samples contain the same set of attributes (also called features, variables or fields). According to the classification of the role of attributes, attributes are divided into conditions properties (also regarded as decision-making attributes) and target attributes (also called type attributes). Thus, a sample is expressed as $(X_1, X_2..., X_m, Y)$. Here X_i is the conditions attributes, Y is the target attribute. The purpose of classification is to find the dependence between the conditions attributes and objectives. This dependence is also named as a classification model or classifier. The classifier is a function. Its input is unknown types of samples, and its output is the type of samples.

Conception Description

From a data analysis point of view, DM is classified into two categories, descriptive DM and predictive DM. Descriptive DM describes the data set in a concise and summary manner and presents interesting general properties of the data. Predictive DM analyzes the data in order to construct a model or set of models, and attempts to predict the behavior of new data sets.

Database usually store large amounts of data in great detail, but clients often like to view sets of summarized data in concise, descriptive terms. Such data descriptions may provide an overall picture of a class of data or distinguish it from

a set of comparative classes. Moreover, users like the ease and flexibility of having data sets described at different levels of granularity and from different angles. Such descriptive DM is called concept description and forms an important component of DM.

The simplest kind of descriptive DM is concept description. A concept usually refers to a collection of data such as frequent buyers, graduate-students, and so on. As a DM task, concept description is not a simple enumeration of the data. Instead, concept description generates descriptions for characterization and comparison of the data. It is sometimes called class description, when the concept to be described refers to a class of objects. Characterization provides a concise and simple summarization of the achieved collection of data, but concept or class comparison provides descriptions comparing two or more collections.

Visualization Processing

Mining data is the process of extracting implied, potential and useful information and knowledge from a large, incomplete, noisy and inconsistent data. Visualization of DM is the integration of visualization technology and DM technology. The human brain can be regarded as a strong and highly parallel processing and reasoning engine. It has a large knowledge base. Visualization of DM can effectively use the human brain. Visualization technology is applied to DM to help people to easily understand the meaning of the data and observe data in a higher level of abstraction. It is convenient for identifying potential models and understanding DM results better.

Visualization of DM is divided into the following areas.

1. **Data Visualization.** Database and the data of a data warehouse are regarded as different in granularity or abstraction levels. They are also looked upon as combinations of different attributes and dimensions. Data are described in a variety of visual ways, such as the box-shaped map, three-dimensional cube, curve surface, data distribution charts, connection plans and so on. This data visualization display gives an overall impression of the data features of a database or data warehouse and allows users to understand where digging commencement point is.
2. **Visualization of DM results.** The visualization of DM results means that the DM results are figured by visual forms, such as a hash map, box-shaped map or other form. The decision tree, association rules, and polarization rules are also described by visualization. This makes it easy for users to understand the results of DM.
3. **Visualization of DM process.** Visualization of the DM process means using the visual process to describe the DM process. In this way, users are able to see which database or data warehouse the data is taken out of, how to extract, integrate, and process, the DM methods used, the storage address of results and the displayed method.
4. **Interactive visualization of DM.** Interactive visualization of DM is a DM process using visualization tools. Users can change the process conditions in interactive ways and observe the impact. This type of exploration allows users to most efficiently search for and discover models, and make the right DM decision-making without the use of automatic DM technology.

METHODS AND TOOLS FOR DMKD

DM includes many technologies and production of multiple different subject fields. This results in a great variety of DM methods and tools. Statistical analysis DM is used to detect exceptional data and explain the marketing rules and business opportunities which are concealed behind the data through statistical models and math models.

However DMKD is quite a different technique from statistical analysis DM. Methods of DMKD include statistical methods, association rules, decision tree, genetic algorithm, rough set theory and support vector machine.

Statistical Method

Traditional statistics supply many kinds of differentiating and regression methods. Common methods are Bayes reasoning, regressing analysis, and variance analysis.

1. Bayes reasoning is the basic tool for modifying probability of the data set after obtaining new information. It is used to solve classification in DM.
2. Regression analysis is utilized to find the optimum model of relationship between input variables and output variables. Linear regression is employed in regression analysis to describe the change tendency of one variable and the relationship with other variables. Logarithm regression is adapted to model the prediction variable set for happening probability of something.
3. Variance analysis is used to analyze and predict performance of the regression line and influence independent variables to ultimate regression. It is one of the most useful tools of DM.

Association Rules

There is common association among things. This is an important content of dialectic. The famous 'beer and diapers' story focuses on the idea that one purpose of business is to find goods which are often purchased at the same time. The relationship between beer and diapers in this example is an association relationship. The process of finding this association is association analysis.

Association analysis was originally a statistics term. It referred to the analysis of association that exists between two or more variables. From a broad point of view, this relationship is a causal relationship or timing relationship.

The more common association is a simple association and timing association in an association relationship. For example, '90 percent of customers who buy bread will also buy milk' is a simple correlation. 'If AT&T stock has risen for two days and DEC stock has not fallen, then the possibility of IBM stock rising in the third day is 7500'. This is a timing association, as its happening is relevant to time.

Association rule mining seeks to find associations in the database. It is one of the issues that is first studied in DM technology. It is also a major research direction of DM. In fact, most of the articles on DM, especially some early research papers and field applications, are focused on the definition of different types of association rules and mining algorithm design. It is possible to conclude that when DM is mentioned, people always think of association rule mining first.

Association rules were proposed by Agrawal, Imielinski and Swami (1993). They were used initially on the mining issue of association rules among the purchases of goods in the customer transactions database of a supermarket. This is the so-called association rules of goods basket data. Agrawal and Srikant proposed the classic algorithm 'Apriori' of mining association rules (1994). It became the most basic algorithms in this area after Agrawal's paper was repeatedly cited. Mining association rules in research papers have now reached thousands of articles, and its application has been extended from the first goods basket data to other data formats. The meaning of rules has also become increasingly diversified. The association rules are classified from different perspectives (Cai W.J., Zhang X.H.& Zhu J.Q., 2001).

According to the variable type of rules, it is divided into Boolean (b-value) association rules and numerical association rules. The data that are dealt with by the b-value association rules

are discrete and classified. Rules are used to show the relationship between these variables. For example,

Bread => Milk

This is a b-value association rule. It shows that the possibility of 'bread' occurring means that the emergence of '"milk"' in a transaction is higher.

According to data dimensions to which the rules have referred, it is classified into one-dimensional association rules and multi-dimensional association rules.

According to data abstract levels in rules, it is divided into single-level and multi-level association rules.

In order to find particularly interesting association rules, semantic constraints are imposed on association rules, such as the left or right of rules must contain certain fields.

The possibility of the simultaneous appearance of data is described by association rules. How to give a quantitative standard to this possibility is a question to be confirmed. In other words, the extentto which such possibility can be considered reasonable. Concepts of supporting extent and believing extent are therefore introduced to define the meaning of association rules statistically.

B-Value of Association Rules Mining Algorithms

After Agrawal (1993) first proposed the problem of mining the association rules among customers' item sets, many researchers made more extensive and deeper study of association rules mining. Their work includes optimizing the original algorithm, such as the introduction of random sampling, data division and parallel processing, improving the efficiency of algorithms, expanding the type of the association rules, proposing numerical association rules, multi-level association rules, fuzzy association rules, sequential pattern, binding and negative association rules, processing distribution algorithms of distribution data and incremental algorithm for incremental mining.

Parallel Association Rules Mining Algorithms

The parallel algorithm is an effective tool for large-scale computing. It employs multiple processors to process data simultaneously. The key issue is coordination in computing, communication and memory use. Because many large-scale databases are dealing with DM and accomplishment of the mining task often requires a longer time, parallel algorithm research in DM has important significance. According to the hardware architecture, the type of information distribution and the type of processor load, parallel association rules mining algorithms are divided into the following categories.

According to the hardware architecture, there are memory distribution, shared memory and hierarchical structure.

According to the type of information distribution, there are three types of algorithms, namely the data parallel distribution algorithm, task parallel distribution algorithm and mixed parallel algorithm.

According to the type of processor load, static algorithms and dynamic algorithms are included.

Numerical Association Rules Mining Algorithm

In numerical association rules mining system, data are first transformed into a binary type and then b-value mining rules are used. For the classification attributes and smaller range attributes numerical type, each value is a variable. Larger range attributes are divided into subsections and then each interval is mapping for a variable. The value in the range is turned to 1, contrary to 0.

Decision Tree

In the classification process, the types of training samples and testing samples are known; classifica-

tion in machine learning is therefore a supervised study. Among the classification methods, decision tree is a common and intuitive rapid classification method.

The decision tree structure usually includes two steps, namely using the training set to generate the decision tree, then carrying out decision tree pruning. The formation of a decision tree is a recursive process starting from root node and from top to bottom. In general, the decision tree is constructed through splitting training samples into subsets. The decision tree pruning is a process that prunes the tree structure and removes redundant branches. The decision tree is used to classify new samples, starting from the root node for testing the attributes of the sample. According to the test results, the next node is determined until it arrives at the leaf node. The category marked by the leaf node is the predicted category of the new sample.

The decision tree structure includes feature selection, node splitting and pruning, and several other steps. Each step is different, and there is a corresponding variety of decision tree algorithms.

The most representative and most common decision tree algorithm is the C4.5 algorithm proposed by Quinlan J.R (1983). Its original thinking can be traced back to the 1950s.

The idea of the decision tree is to structure some assumptions to meet the characteristics of training data. If the noise in data is not considered, the generated decision tree is fully consistent with the training samples, that is, the accuracy of the training data could reach 10000. But given the circumstance of noise, entire consistency will lead to the results of overfitting or overstudying, that is, the decision tree that is consistent with the training samples has relatively low capability to predict the new data. The reason for generating overstudying is that when noise exists in a training set, it is bound to generate some branches to reflect noise in order to be consistent with the training data. These branches not only lead to errors of predicting in the new issue of decision-making, but also increase the complexity of the model.

The simple decision tree not only accelerates the pace of classification, but also helps improve the capacity for exact classification of new data. The more complex decision tree and the more nodes, the less training samples that each node contains, then the lower the representation of the node corresponding to assumptions. Thus may lead to wrong classification. There are two ways to overcome the problem of overstudying. One method is early restraint of tree growth, or stopping growth strategy, the other is tree pruning.

Stopping Growth Strategy

Stopping growth strategy uses certain test conditions to decide whether to continue the division of an impure training subset in the process of generating a decision tree. The stopping growth strategy is also called pre-pruning in some literatures.

Pruning Strategy

The advantage of the stopping growth strategy is that it is rapid, but it easily misses valuable sub-trees. Pruning is another kind of technology to overcome the noise. It is called post-pruning in some literatures. First of all, the decision that is entirely fitting with the training data tree is generated, and then pruning is started from the bottom layer upwards. If the sub-node of one node is deleted, the accuracy of the decision tree (or other evaluation index) is not reduced. This node then becomes a leaf node. The smaller the decision tree is, the easier to understand, and storage and transmission costs are relatively smaller. However, fewer nodes cause a decline in accuracy, so it is necessary to weigh the value of the size of tree against accuracy to prevent over-pruning.

After the decision tree structure is completed, the next step is to estimate its accuracy. New samples with the same distribution from a group

of training sets are usually selected for testing. The test for model is a process to predict the categories of test samples by model and compare the predictable categories with the actual categories to decide whether the predictable categories are consistent with the actual categories. The proportion of samples with the same results in the test sets is the accuracy of the model. When an independent test set is absence, it usually takes the hold-out method or cross validation method to generate training sets and test sets.

The hold-out test method divides the entire data set into two nonoverlapping subsets. The training set is usually 2/3 percent and testing set is 1/3. The decision tree is then structured by using the training set. After that, the decision tree is estimated via the testing set. This testing method is fast, but all the data are not fully used to learning, and it can easily result in overstudying.

Genetic Algorithm

'Natural selection and survival of the fittest' is the basic principle of Darwin's theory of biological evolution. It reveals the law all species always evolve in the direction that is more suitable in the natural world. Only those that adapt to the natural environment can survive and breed future generations. By contrast, those who cannot adapt will be eliminated. After such a lengthy evolution period, a species' adaptability becomes increasingly strong. The characteristics of all species existing on the planet today have come through a long process of evolution. For example, because of the long-term process of scrabbling for food, the deer that has a long neck finds it easier to reach the food that is higher up and thus survives. This is why the giraffe has the characteristic of a long neck.

Genetic mechanisms result in some features of previous generation being apparent in the next or future generations. A person's appearance and walking posture are often similar to his parents, for instance, and this is considered a genetic phenomenon. The material which plays a decisive role in this process is the genetic gene. Parents and offspring cannot be exactly the same and always have some differences, whether greater or fewer. This shows that mutation has occurred in the process of genetic. Mutation is a very important phenomenon of biological evolution and the results of mutation can give birth to a new species.

The genetic algorithm is a random searching method proposed by U.S. scientist John H. Holland in the 1960s. The method can simulate natural selection and the genetic mechanism It is paralleling, non-linear solution and easy to integrate with other models. The genetic algorithm is a global optimization algorithm and is very suitable for large scale searching space optimization problems. Genetic algorithms, evolution strategy and evolutionary programming together constitute the main framework of the evolutionary algorithm, which is the most well-known algorithm of evolutionary computation.

In the process of DM, the problem of finding the optimal solution or a feasible solution is encountered in many cases. Searching in the context of massive data takes a long time to calculate, and heuristic methods such as genetic algorithms can therefore be used in these processes. In fact, hundreds of DM algorithm research papers already address the genetic algorithm optimization problem.

The basic concepts of the genetic algorithm are introduced as follows:

1. **Chromosome:** a bit string formed by a gene, including the biological genetic information.
2. **Coding:** the process of expressing the solution to the bit string. Each of the coded bit string represents one individual - a solution to a problem. The construction of the individual is by coding various parameters of the problem in a particular way, constructing a sub-string, and splicing the string into chromosomes. The most commonly used is binary code. The chromosome length relates

to the accuracy of the problems to solve. If the solution space is real, floating-point encoding is used. Each gene is a variable value, and the length of chromosomes is equivalent to the number of variables. Besides the above two common encoding applications, there are grid code, embedded coding, Gray code and symbols coding. The binary encoding method is simple and easily achieves the evolution operations of cross-evolution and mutation. The Gray code is a deformation of binary code, and similar to binary code. The encoding and decoding operations are comparatively complex, but the local searching ability of the algorithm has been improved greatly. The true value of the variable is used in floating-point coding and the length is much smaller than binary code and Gray code. It needs no repeated data conversion. Its convergence is faster and the quality of the solution is better.

3. **Population:** The group contains a set of individuals, that is, a set of solutions to the problem. The number of individuals in the population is called group size. It is difficult to obtain the optimal solution when the group is too small, and convergence takes a long time when it too big. The general choice range of group size is from 30 to 160.
4. **Fitness:** This is the criterion for evaluating the individual's ability to adapt in the group, which determines whether the solution is good. It is calculated from the evaluation function F. In the genetic algorithm, F is the objective function of solving the problem. The aim is to solve the maximum value or minimum value of the function.
5. **Genetic operator:** The operation of generating a new entity. The common genetic operators are choice, crossover and mutation.
6. **Select:** The individual is directly copied into the next generation of groups. The adaptation of the individual decides the possibility of it being present in the population in the next generation. The higher the individual's fitness is, the higher is the probability of producing future generations. The lower the fitness, the higher is the probability of elimination.
7. **Crossover:** Exchanging the partial gene of two series to produce two new series as the individual of the next generation.
8. **Mutation:** Randomly change the part gene of chromosome, change 0 to one, or 0 to 1.

Rough Set Theory

There are only true and false values in classic logic, but in reality there are many ambiguous values which can hardly be presented simply by true and false values. How to deal with this phenomenon has become a research field, and many logicians and philosophers have been dedicated to research on this vague concept for a long time. As early as 1904, Frege proposed the term vague'. He attributed it to the border, that is, in the whole domain, there are some individuals neither categorized in its subset, nor in the complementary set of the subset.

1965 Zadeh proposed the 'fuzzy set'. Many theoretical computer scientists and logicians are trying to solve Frege's vague concept using this theory. Unfortunately, the fuzzy set is not calculable, which is to say that it did not define the concept of mathematical formula description. It is thus impossible to calculate the number of fuzzy elements, such as subjection function μ in fuzzy set and the arithmetic operators λ in fuzzy logic. In early 1980s and 1990s, Pawlak Z. put forward the rough set according to Frege's border region thinking (1982). He thought that individuals belonged to the boundary line region which could not be confirmed, and the region was defined as the difference set between higher approximate set and lower approximate set. As it has the confirmed mathematical formula description, the number of ambiguous elements can be calculated, that is, the degree of ambiguity between true and false

can be calculated. The main characteristic of rough set theory is that it just reflects the routine of using the rough set approach to deal with the unclear problems. It uses incomplete information or knowledge to handle a number of indistinct phenomena and classifies data based on imprecise results of observation and measurement. After tireless research by many computer scientists and mathematicians, rough set theory has been increasingly improved since the 1980s, especially since the late 1980s and early 1990s in the areas of knowledge where there has been successful application. Compared to other theoretical tools handling uncertainties and fuzzy problem, rough set theory has many unique advantages. After research and development in recent years, it is now employed in information systems analysis, AI, decision support systems, knowledge and data discovering systems, pattern identification and fault diagnosis systems. The rough set concept has similarities with many mathematical tools that were developed to deal with fuzzy and imprecise problems particularly Dempster-Shafer's theory of evidence (Shaw R.D.& Phuoc T.V., 1986). The main difference between them is that Dempster-Shafe uses the function as a major tool, and rough set theory utilizes the collection of lower similar set and higher similar set. Another kind of relationship exists in fuzzy set theory and rough set theory. In short, they are independent ways to incomplete knowledge. There is some relationship among rough set theory identifying analysis, Boolean reasoning methods and decision analysis. One of the main advantages of rough set theory is that it does not need any preparation or additional relevant data information.

In 1980s, many Polish scholars studied rough set theory and its application in depth and unremittingly, and widely analyzed the rough set theory of mathematics and logic. It must be pointed out that rough set theory is not a panacea. The modeling is effective, despite the fact that rough set theory deals incompletely with knowledge, but as a result of the latter, this theory does not contain a mechanism for handling inaccurate or uncertain raw data. Therefore, the simple use of this theory may not effectively describe inaccurate or uncertain practical problems. This means that other ways are necessary. Generally speaking, because evidence theory and fuzzy set theory have ways of dealing with imprecise and uncertain data, it is a natural consideration that rough set theory can be complementary to them.

Support Vector Machine

Support Vector Machine (SVM) is a novel machine learning method that has been developed based on statistical method. Its basis lies in minimizing structure risk to improve the generalization capability of the learning machine. It resolves learning questions effectively because it is easily expended and classifies precisely. SVM is now the substitute method for the training of multi-level perceptron, Radial Basis Function (RBF), neural network and polynomial neural network. In addition, SVM is a convex optimization question and the local optimal solution must be a global optimal solution. These features of SVM are unparalleled by other algorithms including the neural network method. Generally SVM is applied in the classification of DM, regression, exploring unknown things and so on.

APPLICATION OF DM AND KD IN MANUFACTURING SYSTEMS

In general, DM technology is a significant technology in those fields with massive data, such as marketing, risk analysis and identification fraud, as well as in the sports fields. This technology is also being used in manufacturing areas.

Manufacturing intelligence is a very high level of automation of high-tech industries. The various processes feature is one of its characteristics. In the automated manufacturing process, a large amount of data are recorded automati-

cally or semi-automatically, including the basic characteristics of products, working hours and work records of the products in each procedure, manufacturing process parameters and testing data of each procedure Because of the large number of data and increased dimensions, it is difficult for traditional statistical methods to refine the significant guiding knowledge quickly and accurately. Engineers often cannot discover deviations in the manufacturing process quickly and effectively because of the enormous amount of information. If an engineer makes an analysis and judgment of a process situation only by virtue of their professional knowledge and experience, considering the constraints of their own experience, it is bound to result in considerable data wastage and to affect the accuracy and efficiency of the process control. Therefore, the application of DM technology can resolve this conflict to a certain extent. The rate of growth in the amount of resolving information is far greater than the rate of accumulation of knowledge of the problem. Traditional statistical methods can not completely meet the needs of the current information age, and the difference between DM technology and traditional statistical methods is that DM's objective is to find valuable, unknown and potentially useful knowledge for data owners.

DMKD in manufacturing is used in fault diagnosis of parts, resource optimization and analysis of the production process. It is used to ascertain the factors that affect productivity through uncovering some distribution of data that is not normal and exposing the changing circumstances and a variety of factors in the course of manufacturing and assembling to make engineers aware of the scope of the problem so that corrective measures can be taken quickly.

Workshop Scheduling

Workshop scheduling makes adjustments and amendments to workshop task planning according to the issues happening at any moment in the actual processing in the plant. It ensures that workshop manufacturing is accomplished successfully and products are delivered on schedule. Workshop scheduling is an important function for workshop planning and it is supplementary and complementary to manufacturing planning. Workshop scheduling includes two main elements: the ordering of machining workpieces and the distribution of manufacturing resources. Workshop scheduling exists at all stages of the manufacturing process. It is a key issue in ensuring that workshop planning runs effectively in a dynamic real-time environment. Two issues demand particular attention; one is assigning and scheduling the planning task, and the other is distributing planning resources. The machining conflict between delivery and production is a common issue to be resolved. The utilization of hidden sequencing rules can be picked out from historical data to solve the problem of working procedures competing.

Process Parameter Optimization

In the process of the actual machining in the manufacturing system, there are many factors that influence and affect production (Wang, X.Z., 1999). Processing knowledge planning is the core of all data in the optimization of planning parameters. The Computer Aided Process Planning (CAPP) system is usually employed to process data for cutting. After data are extracted, cleaned and transformed, a DM association model is formed. The significance of influence factors is obtained according to analyzing the attributes weighting. Finally, a constituted cutting parameter scheme is predicted via the DM classification model and the scheme is optimized through predicted conclusion.

The scheme of process parameter optimization in cutting factor, cutting tools and surface roughness is described as follows.

1. The attributes weighting method is used to analyze the cutting factor and surface roughness to build a histogram of influence degree about every factor of roughness.
2. The Bayes method is used to make the classification. The model of cutting factor and surface roughness is designed according to history data. Predicted data of this model are input and the outputs are predicted roughness results. The results data which are in the tolerance area are chosen. According to the constraint conditions of these parameters, schemes are filtered and judged, and then the optimized scheme which meets the cutting parameters requirements absolutely is ultimately chosen.

Supply Chain Management

With the development of economic globalization, the advanced supply chain management system gradually becomes the third profit area for companies. The effective use of DM is helpful to companies to build an advantage in information and decision. DM technology is adopted in two modules: the warehouse management subsystem bin choice and the distribution subsystem trails choice.

1. The materials circulation enterprise has plenty of materials circulating and distributing work to deal. An advanced distribution schedule is therefore one of the most important guarantees of enterprise success. In the system distribution management module, the corresponding analysis module is used to choose the optimal transmitting route in solving vehicle scheduling questions in such circumstances as fully loaded or distribution loaded.
2. Bin station chosen is utilized in storehouse management. According to information of goods in storehouse and current bins, the bins are distributed properly. It plays remarkable roles in declining management costs to the materials circulation enterprise.
3. Storage control is also important to supply chain management and is an issue that every enterprise which has storage production requirements must face. All productions are first classified, and then the association results are obtained via the association rules. Finally, according to above model, database is mined to get the decision tree.

Quality Control

Quality Control (QC) is an important step in manufacturing and QC methods are a powerful issue in manufacturing research.

The focus of study is on methods based on DM; many researchers propose a variety of methods, such as QC based on dynamic DM, applying DM to the control map and using various mining algorithms, such as rough sets, or association rules mining algorithms to improve the efficiency of QC. However, with the development of a new manufacturing model, in the circumstance of no mining experts, ontology theory and methods are employed to use data mining to achieve better QC.

The basic idea of quality modeling based on DM is to firstly select the historical data of the modeling object; after pretreatment, control rules are formed through the extraction of hidden and relevant knowledge of quality. According to DM steps to process quality data, quality model is accomplished. Ontology is used to overcome the limitations that experts present in the traditional DM process. It is possible for non-professional people to easily complete the excavation of the modeling process.

CASE STUDY

In this section, we will present a case study about application of DM in process parameter optimization.

Background and Motivation

In the actual machining process of flexible manufacturing system, there are often many factors resulting in the impact on manufacturing results. It is very difficult to give precise mathematical model by using traditional reasoning methods. The classification analysis method combining with data mining is proposed in this section, and by using them to establish classification about the roughness results of cutting experiment, the targets are locked on the smaller local scope to find solution, making it possible to optimize the cutting parameters.

DM Model of Process Parameter Optimization System

Data Mining Model of the Process Parameter Optimization

In the process parameter optimization, process design knowledge is the core of all the data. For the cutting process data in the CAPP system, after they are extracted, cleared and converted, through setting up relational model of data mining and using the property-rights analysis, the importance of influencing factors is obtained. By using classification model of data mining, the established cutting parameter project is predicted, and the project is optimized referring to forecast results.

Data mining is a cross of many subject disciplines, using technologies in many subject disciplines such as statistics, computer, and mathematics and so on. It is not only a simple search or a query call facing to the particular database, but also it is required to do data statistical analysis, synthesis and reasoning microscopically and macroscopically, in order to guide the solution of practical problems and to attempt to find linkages between events, even to predict future activities by using the existing data. The popular data mining methods include commonly classification, clustering, dimension reduction, pattern recognition, visualization, decision tree, genetic algorithm, uncertainty process and so on, through the seamless integration between database and data mining algorithm, the data mining model is built. It is shown in Figure 2.

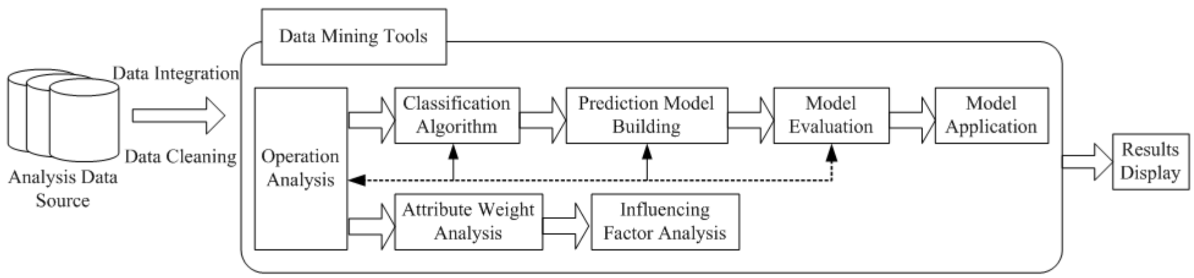
Data mining algorithms are consisted of attribute weight analysis algorithm and classification algorithm. The results are predicted through building classification model by adopting data mining tools. The results are displayed to the users through visualization technology.

The Application of Data Mining Technology in Process Parameter Optimization

For turning, when the external conditions (tool, work-piece material, machine tools etc.) are under certain circumstances, the general control parameters are cutting speed v, feed rate f and cutting depth a_p. For the process, as the cutting depth is often pre-determined in accordance with work-piece materials and process sizes, optimization is made mainly for cutting speed and feed rate, when the operator do the actual processing. In order to facilitate the description, the analysis is done only for the three cutting elements, tool abrasion loss and surface roughness. Specific programs are as follows:

1. The relationship between cutting factors and surface roughness is analyzed by using attribute weight analysis, the impacts on surface roughness referring to the various factors are accessed, and the results are displayed using histogram;
2. Classification is done by using Bayesian algorithm, and the model between cutting

Figure 2. Data mining model of process parameter optimization



factors and surface roughness is built by using historical data, the forecast data sets are considered as inputs, the predicted roughness results are obtained.

Select all structural data records which are within a certain error scope against the scheduled roughness requirements, these programs are filtered and assessed in accordance with the constraints of the physical meanings of these parameters, so that the optimization program of cutting parameters which meets the requirements can be selected. Optimization process of cutting parameters is shown in Figure 3.

Attribute weight analysis using MDL (Minimum Description Length) principle algorithm, the relationship between cutting factors and surface roughness is analyzed by using attribute weight analysis, the impacts on surface roughness referring to the various factors are obtained, and the results are showed in sequence.

The measured process state parameters include: work-piece speed n_w, feed rate s, cutting depth t, as well as the surface roughness R_a.

The optimization target of data analysis is to find the optimal cutting speed and feed rate after analyzing the initial data, to meet the requirement that roughness reaches 1.6 um. The selected sample data is 10,000 in total,

In the process of data analysis with the method of attribute weight analysis algorithm and classification analysis algorithm, the surface roughness in the sample data needs doing discrete process. It can be separated into three intervals according to roughness grades: <1.6 μm, [1.6 μm, 3.2 μm), [3.2 μm, 6.3 μm). After the transformation of data samples, finally a total of more than 10,000 data for mining is obtained and kept in the database which meets the quality requirements.

The weight values shown in Table 1 are the probability statistical values of impacting roughness corresponding to different attributes. We can see that the importance of various factors is arranged according to the following sequence: the work-piece speed n_w> feed rate s> cutting depth t. Among them, the attribute value of work-piece speed n_w is 0.68, which is the most important factor for the impact on surface roughness, therefore, in the process of cutting quality check, when disqualification condition of surface process roughness is found, the reasons for work-piece speed can be considered in prior according to the results of system analysis.

Bayesian algorithm is chosen by classification algorithm to do the classification analysis, Bayesian algorithm is a kind of algorithms in which probability statistics knowledge is used to do the classification. Each sample should belong to a particular classification, assuming there are m-categories $C_1, C_2, \ldots, C_m$. The samples unknown classification are classified to C_i, only when $P(c_i|x)>P(c_j|x)(1\leq j\leq m, j\neq i$, in which x is a sample unknown classification, c_i represents classification c_j, $P(c_j|x)$ represents the probability that x belongs

Figure 3. Flowchart of cutting parameters optimization

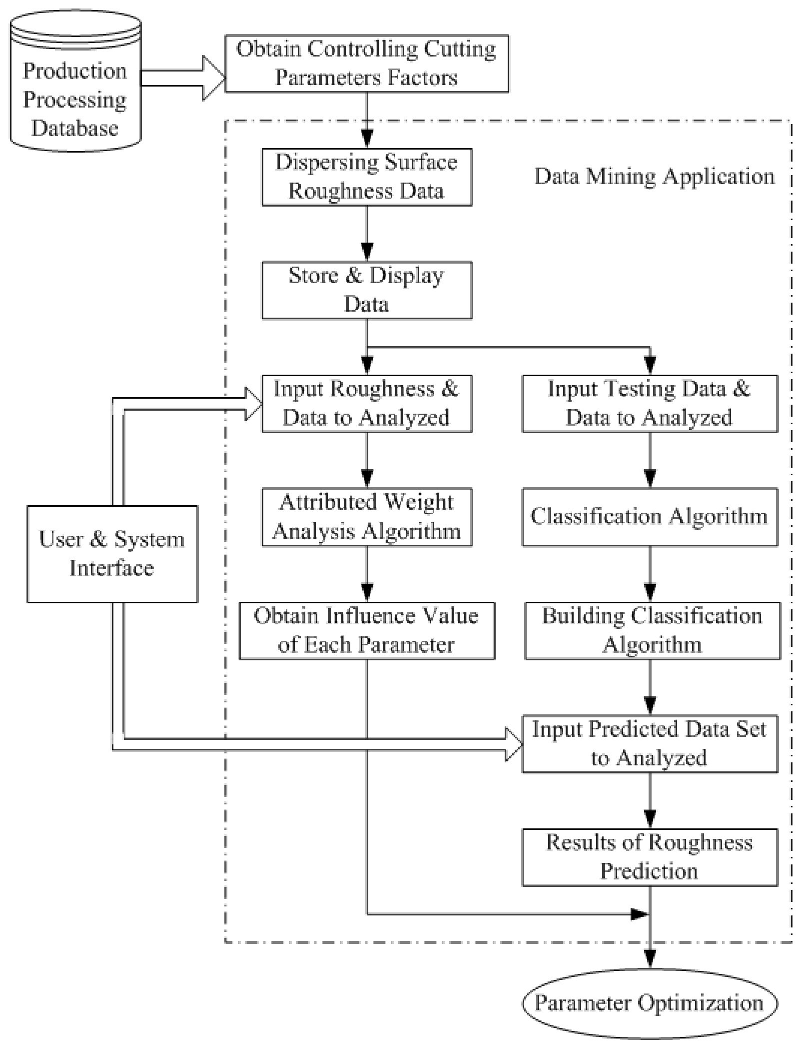


Table 1. Results of attribute weight analysis

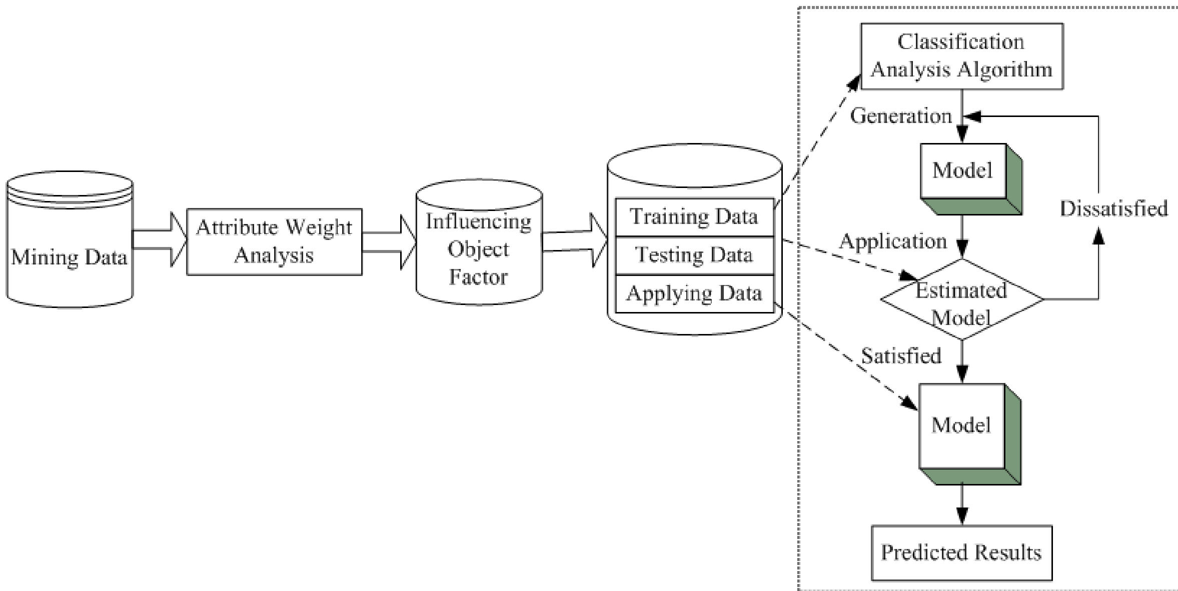
Name of attribute	Weight	Rank
work-piece speed n_w	0.68	1
feed rate s	0.54	2
cutting depth t	0.45	3

to classification c_j; at the same time, according to Bayes theorem $P(C_i \mid X) = \frac{P(X \mid C_i)P(C_i)}{P(X)}$, the determined condition can be simplified that the value for i makes formula 4.1 has the greatest value; namely the samples unknown classification are putted into the formula 4.1 to determine that the i ($1 \le i \le m$) takes which value can make $P(X|C_i)P(C_i)$ the largest value, then the sample x belongs to the classification C_i.

$$P(X \mid C_i)P(C_i) = \prod_{k=1}^{n}[P(X_k \mid C_i)]P(C_i) \qquad (1)$$

The model forecast accuracy can be adjusted by adjusting the entrance parameter values of Bayes-

Figure 4. Data processing



ian function, repeat the first step until satisfactory accuracy is obtained, the entrance parameter is the minimum man-made set reference value of the $P(X_k|C_i)$ in formula 4.1, when the $P(X_k|C_i)$ is less than the set entrance parameter, the value $P(X_k|C_i)$ is substituted by the set entrance parameter value, the entrance parameter is usually set combining with the analysis results of attribute weight values, the entrance parameter varies between 0-1.

For the data in classification analysis, it is divided into training data and test data in accordance with a certain proportion. It can not only ensure that there are sufficient data samples to generate the data mining model of process parameter optimization, but also be able to test the validity of detection model. If the results of the detection model are not satisfactory, classification model can be re-generated by adjusting the relevant parameters, the process is repeated until such a model is satisfactory. The specific data processing is showed in Figure 4.

By using Bayesian algorithm to classify, the predicted results of roughness are compared with the roughness of the prior same product, to determine whether there is a need to optimize the roughness. By choosing the data records that meet the scheduled requirement - their roughness is within a certain error margin, these process parameters are filtered and assessed in accordance with the constraints of the physical meanings of these parameters, integrating the rules of association knowledge discovery, the optimization program of cutting parameters which meets the requirements is selected.

Select all structural data records which are within a certain error scope against the scheduled roughness requirements, these programs are filtered and assessed in accordance with the constraints of the physical meanings of these parameters, so that the optimization program of cutting parameters which meets the requirements can be selected.

By using Bayesian algorithm to classify, the imported historical data sets are considered as the analysis objects and the surface roughness Ra of objective properties, the model of classification

algorithm is established, and the accuracy of the model is verified by using test data. Through calculating, the obtained system accuracy rate is 90.87%, which gives the test number between the actual values and the expectation values. If the accuracy of the test model is not satisfactory, classification model can be re-generated by adjusting the entrance parameter values of Bayesian function, namely the $P(X_k|C_i)$ in Equation (1), and the process is repeated until such a model is satisfactory. After the establishment of the forecasted model, the cutting parameter values which need to be optimized can be predicted.

Since the uses of Bayesian algorithm have some limitations, there is a certain degree of uncertainty for the classification prediction, you can choose according to the actual situation. When the cutting parameter program in which the predicted roughness is more than the requirement value can meet the requirements in practice, it also can be regarded as an optimization program to choose.

Conclusion

By applying the property weight analysis in data mining technology, the importance of cutting parameters which have impacts on the process surface roughness can be putted in order, so when disqualification condition of surface process roughness is found, the factor which has the highest weight value in cutting parameters should be considered in prior according to the results of system analysis. In addition, by using the classification analysis to generate predictive model, the processing parameters are predicted, the optimal combination of processing parameters can be predicted, in the circumstance that the processing requirements are determined. It is propitious to improve the machining accuracy in the actual processing, so that the product quality and production efficiency can be enhanced.

SUMMARY

DMKD is a process of identifying effective, innovative, potentially useful and finally understandable models. DM is one of steps in KDD that generates specific mode through the specific algorithm in an acceptable computational efficiency. KD is an entire process that includes the steps of data choice, data preprocessing, data transformation, DM, and pattern evaluation, to eventually acquire knowledge.

At present, although DM technology has progressed to a certain level of application, and achieved remarkable success, there are still many unresolved issues, such as data pre-processing, mining algorithms, pattern recognition and interpretation, visualization, and so on. For the manufacturing process, the most crucial question DM poses is how to combine temporal and spatial characteristics of manufacturing data, and express mined knowledge. The expression is the space-time knowledge and interpretation of the expression mechanism (Xingquan Zhu, Ian Davidson, 2007). With in-depth research into DM technology, is hoped that DM can be applied in more extensive areas of application, and that more significant results can be achieved.

The development trend of DM is expected to proceed as follows.

1. **Visualization of DM:** an effective way to find knowledge from large amounts of data. Visualization makes the process of DM comprehensible by the user, and also facilitates human-computer interaction in the course of DM.
2. **Multimedia DM:** in processing DM from a large amount of text, graphics, data, video image data, audio data and integrated multimedia data, an implied and valuable model is discovered through semantic analysis and seeing and hearing features. Its disposed data

differ from those in traditional DM methods. Data in traditional DM processing are records and items of table format in the database, in other words, structured data. The data treated in multimedia DM are unstructured data.

REFERENCES

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. In ACM SIGMOD, (pp. 207-216).

Agrawal, R., & Srikant, R. (1994). Fast Algorithm for Mining Association Rules. In J. B. Bocca, M. Jarke & C. Zaniolo, (eds.), *Proceedings of 20th Internation Conference Very Large Data Bases (VLDB)*, (pp. 487-499). San Francisco: Morgan Kaufmann Press.

Cai, W. J., Zhang, X. H., & Zhu, J. Q. (2001). Summarization of Association Rules Mining [J]. *Chinese Computer Engineering*, *5*, 31–33.

Fayyad, U. (1998). Mining Databases: Towards Algorithm for KD. *IEEE Bulletin of the Technical Committee on Data Engineering*, *21*(1), 39–48.

Fayyad, U., & Piatesky-Shapiro, G. (1996). The KDD Process for Extracting Useful Knowledge form Volumes of Data. *Communications of the ACM*, *39*(11), 27–35. doi:10.1145/240455.240464

Frawley, W., Piatesky-Shapiro, G., & Matheus, C. (1991). *Knowledge Discovery in Databases* [M]. Cambridge, MA: AAAI/MIT Press.

Han, J. W., & Kamber, M. (2001). *DM: Concepts and Techniques* (pp. 2–4). San Francisco: Morgan Kaufmann.

Hand, D., Mannila, H., & Smyth, P. (2002). *Principle of DM* (pp. 1–2). Cambridge, MA: MIT Press.

Pawlak, Z.(1982). Rough Sets. *International Journal of Computer and Information Sciences,* (11), 341-356.

Pawlak, Z. (1990). Rough Sets Theoretical Aspects of Reasoning about Data. *Nowowiejska 15*(19).

Quinlan, J. R. (1983). Learning Efficient Classification Procedures and Their Applications to Chess and Games . In Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.), *Machine Learning: An AI Approach* (pp. 463–482). Palo, CA: Tioga.

Shaw, R. D., & Phuoc, T. V. (1986). Antigenic Mapping of the Surface Proteins of Rhesus Rotavirus. *Virology*, *155*, 434–451. doi:10.1016/0042-6822(86)90205-9

Wang, X. Z. (1999). *Data Mining and Knowledge Discovery for Process Monitoring and Control*. London: Springer.

Witten, I. H., & Frank, E. (2000). *DM: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco: Morgan Kaufmann.

Woksem, J. (2001). Treads on Database: Reasoning and Mining. *IEEE Transactions on Knowledge and Data Engineering*, *13*(3), 426–438. doi:10.1109/69.929900

Zhu, X., & Davidson, I. (2007). *Knowledge Discovery and Data Mining: Challenges and Realities*. Hershey, PA: IGI Global.

Chapter 5
Computational Intelligence

In the 1990s, a new paradigm of science characterized by uncertainty, nonlinearity, and irreversibility and tackling complex problems was generally recognized by the academic community. In this new paradigm, traditional analytical methods are ineffectual, and there is recognition of the need to explore new methods to solve the more flexible, more robust system problems. In 1994 the first Computational Intelligence Conference in Orlando, Florida, US, first combined three different areas, smart neural networks, fuzzy systems and genetic algorithms, not only because the three have many similarities, but also because a properly combined system of the three is more effective than a system generated by one single technical field. Various theories and approaches of computational intelligence including *neural computing*, *fuzzy computing* and *evolutional computing* are comprehensively introduced in this chapter.

DOI: 10.4018/978-1-60566-864-2.ch005

INTRODUCTION

Computational Intelligence is generally considered to incorporate Neural Networks (NN), Fuzzy Systems (FS) (Dong, 2008) and Evolutionary Computation (EC) (Cacciola, 2008)) as its three main areas. Its positive significance is to promote the comprehensive integration of various kinds of smart theories, models and methods which are based on the combination of calculation and symbols, in order to facilitate the development of intelligent behavior which is more advanced on thinking, and more powerful. Computational intelligence is capable of solving more complex systems problems.

The current key issues of computational intelligence are divided into three main areas:

1. **Explain the calculation theory of calculation intelligence.** The most typical is evolutionary computation. Researchers have been keen to

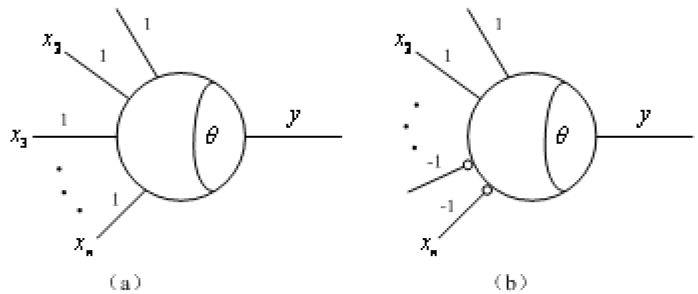
explain the calculation theory of GA, and to establish the precise mathematical theoretical basis, but this research is still in the exploratory stage.

2. **Put forward new calculation models (Christine, 2002).** There are many neural network models, but most of them offer a one-sided understanding of the biological nervous system. The establishment of a closer understanding of the human neural network model is expected based on the further exploration of the mysteries of human intelligence. Biological evolution seems simple, but GA is unable to fully understand and imitate it, so the improvement of GA (Cantu, 2000) is needed to promote its performance. Researchers have noted that in the course of evolution DNA fragments are explained through RNA (Ribonucleic Acid), amino acids, and proteins, unlike the direct explanation given by GA. At the same time, there are mechanisms in biological systems such as the ant colony in search of food, or the body's immune system's ability to resist the effects of viruses, which are used for reference by evolution calculation to establish the structure of a new model .
3. **Combine various calculation methods to create a new calculation model (ClercM, 2000).** The earliest neural networks can achieve good computing performance by applying the idea of simulated annealing to generate a random neural network model. The optimization of the parameters of the neural network structure is by virtue of GA calculation. GA is easily combined with other methods to form hybrid genetic algorithms (HGA), such as in the choice of operation in the simulated annealing method, or in the decoder algorithm in other local searching operations. The integration of fuzzy systems and artificial neural networks allows wide coverage of the application, so the fuzzy neural network is also a burning issue in computational intelligence.

Computational intelligence is widely used in many areas, covering all areas of AI, such as knowledge acquisition, associative memory, speech recognition, image analysis and so on, as well as penetrating into other areas (Clec, 2002). With the development of complexity of calculating theory, intelligence shows a very strong capacity for solving complex planning problems, for example, the complex combinatorial optimization problem solving (TSP), the multi-function complex calculation of the maximum (Katayama, 2004). Computational intelligence methods and related technologies, such as the Simulated Annealing (SA) algorithm and Tabu Search (TS) constitute the basis of modern optimization technology. The applications in the field of optimization technology in turn promote the progress of computational intelligence technology.

Computational intelligence is widely applied in manufacturing (Antonio, 2000). The essential foundation of intelligent control is fuzzy control and neural network control technology (Chen, 2004). In recent years, with the progress of real-time genetic algorithms, the GA algorithm has been widely used in the field of intelligent control (Ritter, 2007). Researchers are used to applying computational intelligence methods to resolve many complex decision-making problems in manufacturing, such as inventory planning, shop layout, production process scheduling and control (Srivastava, 2008), which are of the Non-deterministic Polynomial Completeness (NPC). The study of these applications has brought about many valuable algorithm models, which have an important role in the theoretical study of computational intelligence.

Figure 1. (a) M-P artificial neural model, (b) its early forms



ARTIFICIAL NEURAL NETWORKS

Neuron Model

Biological neurons building up neural networks have the following characteristics: (1) Two states. One is excitement and the other is curb work. (2) Threshold effect. Above a certain threshold, the neurons excite. (3) Multi-input and single-output. Neuron cell bodies receive lot of input through the tree-shape protrusion from other neurons, and export single-output through the field axon. (4) Space and time superposition. Numbers of neurons input come into effect in space and time by way of stacking. (5) Plasticity connector. The strength of synaptic connections is adjusted. In 1943, McCulloch and Pitts defined a simple artificial neural model, known as model MP, the inspiration for biological neurons. The model is shown in Figure 1 (a) below. Early input and output are only from (0,1) value, and the weight takes the value 1 (excited) or 1 (inhibited), as shown in Figure 1 (b).

For later developed general artificial neural models, the following formulas are used

$$X=[x_1,x_2...,x_n] \quad (1)$$

$$W=[w_1,w_2,...,w_n] \quad (2)$$

Respectively represent one group of the input and the corresponding weights of plasticity of the simulation synaptic, and θ is the threshold, then that the output of neurons is denoted as

$$Y = f\left(\sum_{i=1}^{n} w_i x_i - \theta\right) \quad (3)$$

$f(\cdot)$from units such as the function of

$$f\left(u\right) = \begin{cases} 1, u \geq 0 \\ 0, u \prec 0 \end{cases} \quad (4)$$

Although the MP model is simple, it basically reflects the first four characteristics of the biological neurons' five characteristics. The later more complex weight-adjustable artificial neural network is also based this model.

Three important factors of the artificial neural network are: node (that is, artificial neuron) input-output characteristics; network topology; learning algorithms of weight-determining. In fact, some function approximation problems are equivalent to the artificial neural network with certain weight.

Classification of the Artificial Neural Network

The reason why the human brain has the highly intelligent capacities of memorizing, associating, deducing and judging is that numerous neurons are connected as a network in a certain way to work together, and the intensity of the synaptic

junction between each part is adjusted according to certain rules. The function of ANN simulating the biological nervous system is determined by two factors. One is the structure of the network, or in other words the connection of the network. The other is the learning and operating rules of the network, namely the adjusting rules of the connection weights in the network. Though ANN try hard to simulate the structure and function of the biological nervous system, it still has a long way to go in respect to the number of the neurons, the complication of the network structure and the intelligence of the network.

Single-Layer Feed Forward Neural Network

The input layer cannot be regarded as a layer because it only performs the functions of allocating the input signals and transmitting the information. The output layer is used to process information as well as output the result.

Multi-Layer Feed Forward Neural Network

The research of physiologist reveals that the human brain cortex contains 3 to 6 layers of nerve cells. That means there is hierarchical structure in the human brain cortex. Simulating the biological nervous system, ANN is arranged by hierarchy. Multi-layer network is the result of cascade connections of single-layer networks, which means that the input of neurons of the next layer is the output of the former layer.

The neurons of the input layer assume the tasks of receiving the input information of the external environment and transmitting it to the neurons of the middle hidden layers. Every hidden layer is the internal signal-processing layer in charge of the transformation of the information. The last hidden layer transmits the information to the output layer. Then the output layer further process the information and output the final result to the external environment. Thus one input-to-output information process is done. If the network is a multi-layer one made up of output layers, the transition function between each layer must be non-linear, or else the multi-layer network can be only regarded as single-layer network because it has no more mapping and storing capacity than the single-layer network.

The weight matrixes of the first, second and third layers are respectively W1, W2 and W3, $i=1,2,\dots,n$; $k=1,2,\dots,p$; $l=1,2,\dots,q$; $j=1,2,\dots,m$. The input and output of that three layers are respectively Γ_1, Γ_2, Γ_3, then

The output of the first layer is: $Y_1=\Gamma_1 W_1 X$

The output of the second layer is: $Y_2=\Gamma_2 W_2 Y_1=\Gamma_2 W_2(\Gamma_1 W_1)=\Gamma_1\Gamma_2 W_1 W_2 X$

The output of the third layer is: $Y_3=\Gamma_3 W_3 Y_2=\Gamma_1\Gamma_2\Gamma_3 W_1 W_2 W_3 X=\Gamma WX$

Where $\Gamma=\Gamma_1\Gamma_2\Gamma_3$; $W=W_1W_2W_3$.

Interconnected Feed Forward Neural Network

The mechanism of widthwise restraining or activating between the neurons in the same layer can be realized by the interconnection of the neurons with the same layer. With this widthwise mechanism, the number of the active neurons at one time can be limited, or the neurons of each layer can be divided into several groups and each can work as a whole. For example, the neurons of the maximized output can be selected out, meanwhile the other are restrained to the state of no output.

Feed Forward Neural Network with Feedback

In feedback neural network which is based on feed forward neural network, the information of the output layer is fed back to the input layer or the formal layer through the connection weight. The output of this network is not only related to the current input but also to the formal output, which has the ability of memory for short time.

Interconnected Joint Neural Network

Interconnected joint network can be divided into two types: the wholly interconnected network and the partially interconnected network. In the former, every neuron is connected to the other, such as Hopfield network and Boltzmann machine. In the interconnected joint network, signals are repeatedly transmitted between neurons, so the network is in constantly dynamic change and only through several times changes can it achieve balance. Because of the different characteristics of transformation functions of neurons or networks, the networks may develop into periodic oscillation or other states such as fuzzy.

Basic Model of the Artificial Neural Network

The existing neural network has dozens of models. With the further development of theoretical study and application, new models will continue to emerge. This section introduces some common neural network models, including BP(Error Back Propagation) model, Hopfield model, Boltzmann model and ART(Adaptive Resonance Theory) model.

BP Model

In 1985, Rumenlhart (USA) proposed a back-propagation neural network algorithm, BP model, which broke through the restrictions of the two-tier network, joined the implied nodes, and provided an effective learning algorithm. According to the value of error of learning, the connected weight of each floor from behind to front was amended so that the learning errors of the network achieved a minimum.

The learning process of the algorithm consists of the positive transmission and the reverse transmission. In the process of the positive transmission, input information is processed by layers from the input layer through hidden layers and transmitted to the output layer, and every layer of neurons in each state only influences the next layer. If the output layer can not receive the expected output, then it is spread into reverse transmission, and the error signal is moved back to the original connection along the pathway, by amending the weight of each lever so as to receive the smallest error signal. In theory, the node role function of BP network is required to be a differentiable function, the most commonly used being Sigmoid, hyperbolic, the sine function, and S-type function.

$$f\left(x\right) = \frac{1}{1 + e^{-x}} \tag{5}$$

$$f\left(x\right) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{6}$$

$$f(x)=\sin x \tag{7}$$

Although the BP model in all its aspects has its own significance, it raises the following questions:

1. From the mathematical perspective it is a nonlinear optimization problem, and it is inevitable that there is a local minimum, which is also often encountered by the gradient.
2. The learning algorithm convergence is very slow, usually several hundred or even a 1,000-step iteration.
3. The selection of the number of the hidden network nodes has no theoretical guidance, but is only based on experience and a series of tests to select.
4. The addition of new samples have an effect on the samples that have already completed study; the number of the characteristics describing each input sample is required to be the same.

Hopfield Model

In 1984, Hopfield put forward a new network model – now known as the Hopfield model. As

early as 1982, Hopfield had proposed a discrete stochastic model, and the random discrete and continuous-time models of the Hopfield model are highly relevant. This section only discusses the continuous-time Hopfield model, mainly because it is easy to implement in hardware.

In Hopfield a network, with a group of non-linear dynamic formulae to describe the output of each operational amplifier (neurons) in the network, the input of operational amplifier is described as

$$\begin{cases} C_i\left(\frac{du_i}{dt}\right) = \sum_j T_{ij}u_j - \frac{u_i}{\tau_i} + I_i \\ v_i = g_i \bullet \left(u\right) \end{cases} \tag{8}$$

C_i is the overall input capacitance of the operational amplifier, T_{ij} is the connected weight from the j-operational amplifier to the i-operation amplifier input, $v_i$is the output of i-operational amplifier, τ_i is the value of resistance determined by $g_i(\cdot)$,$g_i(\cdot)$is the S-type transfer functions of the i-operational amplifier (assuming operational amplifier response time is ignored), I_i is on behalf of the external input of i-operational amplifier. To make T_{ij}a positive or negative, use reversed or overturned operational amplifiers, the values of T_{ij} are $1/R_{ij}$, $R_{ij}$is the resistance from the input of i-operational amplifier to the output of the j-operational amplifier. The value of $\tau_i$is determined by the following formula:

$$\frac{1}{\tau_i} = \frac{1}{\rho_i} + \sum_j \frac{1}{R_{ij}} \tag{9}$$

Here ρ_i is the input resistance of the operational amplifier.

In order to ascertain the operation of the Hopfield model, the output value of the i-operational amplifier is considered as the i-component of the N-dimensional state vector, in which N is the number of neurons in the network, and the state of vector and Equation (8) move in the direction of the mobile space. To select the internal strength of connections only, the value of state-space flow and operating characteristics of the network are determined. For the neural network system identified by (8), the following conclusions are reached:

The energy function of the system is monotone decreasing along its balance track; if the balance point is stable, it is asymptotically stable; the asymptotically stable equilibrium point is the local minimum of E; systems process the information by way of a large-scale nonlinear continuous-time parallel. The time of calculation is the time that the system requires to achieve stability.

Boltzmann Model

In 1985, using statistical physics concepts and methods, Hinton proposed a new neural network model. The probability distribution of the model used the Boltzmann distribution of statistical physics, and is therefore, called the Boltzmann Machine.

A support network is set up by N neurons, in which each neuron has its own state, and is expressed as:

x_i-0 or 1,i=1,2,...,n

Assume that the bond strength between neurons is symmetrical, that is the weight

$_{wij}$=w_{ji}, i,j=1,2,...,n

In the network, the changes to the sum of the input neurons will lead to the status update of the neurons. This update in various units is a synchronization, which is represented by probability distribution methods. According to the Boltzmann principle, when i-neurons are in the status update, the probability of the new status which is 1 is:

$$P\left(x_i = 1\right) = \frac{1}{1 + e^{-\Delta E_i / T}} \qquad (10)$$

where the T is called the 'temperature' of the network, and is positive. ΔE_i is the energy difference of the systems when the state of i-neurons is 1 and 0.

$$\Delta E_i = \sum_j w_{ij} \cdot x_j - \theta_i \qquad (11)$$

when the input increases, the probability of the status of 1 will increase. At the same time, the changes of the curve have something to do with the temperature. When the temperature T becomes high, the curve's change is relatively gentle. When the T is close to 0, the curve of probability P is close to step function.

The above is similar to the changes of particles in a group in a thermodynamic system. When these particles in a group approach a heat source, they will eventually reach a certain heat balance, and then the state probability of the whole group will meet Boltzmann distribution:

$$\frac{P_\alpha}{P_\beta} = e^{-\left(E_\alpha - E_\beta\right) / T} \qquad (12)$$

Here P_α represents the probability of network with state α.E_α represents the energy of the state α. The above formula shows that when the temperature T = 1, the logarithm difference of the emergence probability of the two states in the network is equivalent to the energy difference of the two states.

In this way, when the input neurons of the network are fixed in a particular input state, the network should be able to identify the minimum energy structure which is most suitable for specific input and a certain temperature is required to reach thermal equilibrium so that the network can reach a rough energy minimal. To achieve an overall energy network minimum, at the beginning, a higher temperature is needed, then using simulated annealing technology, the temperature is gradually reduced to find a better energy minimum.

ART Model

ART Model is a non-guided learning rule which relies on a self-organized way to respond to input mode sequence, and classifies the input mode by distinguishing the input information.

ART usually has three layers, of which the output layer and input layer are known as the short-term memory layer, and the middle layer is known as long-term memory. There exists a feedback vector from the output layer to the input layer called the top-down stabilization mechanism. The long-term memory layer remembers what has been learned, and the learned mode is stored in the long-term memory storage layer in the way of weight. The hort-term memory deals with the current input mode, the classification of the mode and the desired pattern. The top-down stabilization mechanism ensures the firmness of the memory in the long-term memory layer. Whenever a model is input to the system, the desired mode is picked out from the long-term memory for comparison. If the mode of input is similar to the desired mode, the input mode will be assigned to the corresponding category. If there are no similarities, a new category represented by the input vector will be formed and stored for the first time as a new class in the memory preservation.

The non-guided learning method of ART is known as 'competitive learning'. This technology enables neurons with a limited number of active duties to compete with each other. The simplest competitive learning is the so-called 'winner take all' rule, in which the neurons with the maximum input are activated and the rest of the neurons are restrained.

The ART Rule is a learning rule which is relatively close to the human brain's method of processing information. It is 'real-time learning', that is, study and work are inseparable. Stable and

reliable results are achieved via this learning. This study is one of autonomy and self-organization, a learning process without the guidance of teachers, so it is a non-supervised learning. This study could form the 'focus' of the state, which is having the choice of study, and which has a high learning efficiency. This study could also completely avoid local minimum points.

Learning Rules of Artificial Neural Networks

Artificial neural network simulating the activity of human brain has three categories of learning rules: supervised learning, unsupervised learning and reinforce learning.

In the supervised learning, the errors are minimized by adjusting the weights of the network according to the errors between the network output and the objective output. In the unsupervised learning, there's no output but input in the learning data set. The feature of input data is automatically extracted by the neural network during the learning, and is divided into several types. The trained network can recognize the new input data besides the used learning data set, and accordingly get different output. The reinforce learning is characterized by adapting to the environment because of inputting the feedback of the environment. In this method, the environment is mapped to the action of the system, and the objective is to maximize the accumulated reward value acquired from the environment. It's different from the supervised learning which informs the system which action to be taken using positive or negative examples, whereas taking the trial-and-error method to discover the optimization action policy. The regular learning rules of neural network are as follows.

Hebb Rule

In the late 1940s, Donald Hebb first proposed a neural network learning algorithm, called the Hebb Rule. Based on this rule, a variety of learning algorithms have emerged. The Hebb Rule was preceded by Hebb's Law, which became its foundation.

Hebb's Law is the basic law in psychology of linked learning, which is also the basic law of the artificial neural network. Hebb's Law holds that all associative (long-term) memory is stored in the synaptic connections of the central nervous system, and all learning means changes in the synaptic strength of coupling. The Hebb Rule is founded on Hebb law. Assume that the strength (weights) of connections from j-neurons to the i-neurons is w_{ij}.Δw_{ij} is the change of the weight. a_i^u and a_j^u represent activation state of i-neurons and j-neurons on condition that $P(p=1,2,...,p)$ is the input samples. The Hebb Rule is denoted as:

$$\Delta w_{ij}=\eta\cdot a_i\cdot a_j \tag{13}$$

$$w_{ij} = \eta\sum_{u=0}^{p-1} a_i^u \bullet a_j^u \tag{14}$$

Here η is the speed factor of learning.

The superiority of the Hebb learning rule is reflected in the correlation of the incentive value of i-neurons and j-neurons, which can produce useful association studies. For specific neurons, it will tend to encourage those related activation neurons when it is in the activation state.

Delta Rule

The most popular neural network learning algorithm is the Delta Rule which is an extension of Hebb rule and was put forward by Widrow and Hoff in the 1960s. Widrow and Hoff compared the input and output using adaptive linear components and generated real and expected output through changing the weights, the main purpose of which was to reduce the standard deviation signal. In order to reduce the error between the real output value and the expected output value, the weights need to be adjusted. The rule of adjusting weights

is the Delta Rule, which is written in the following formula:

$$\Delta w_j=\beta(d(p)-y(p))\cdot I_j(p)=\beta\cdot\delta(p)I_j(p) \quad (15)$$

Δw_j represents the change of J-weight. $d(p)$ is the expected output of the p-input. $y(p)$ is the estimated output of the p-input. $I_j(p)$ is the j-component of the p-input. β is the learning rate.

The Delta Rule is the gradient; each derivative of weight is in inverse proportion with the change of weight.

Kohonen Rule

The Kohonen Rule also adopt the concept of competitive learning, and is a non-supervisory study rule.

Suppose i-neurons gets a lot of input x_{ij} from other neurons, these inputs constitute a vector $x_i=\{x_{i1},x_{i2},...x_{in}\}$, the corresponding weight of each input x_{ij} is w_{ij}, each of the weight can also be seen as a vector $w_i=\{w_{i1},w_{i2},...w_{in}\}$. When the neurons connect function is a summation function, the result is the dot metrix of the input vector and the weight vector.: $x_i\cdot w_i$, and it is the measure of the distance between the two vectors.

The Kohonen Rule also uses the 'winner take all' strategy, and the winners are the neurons whose weight vector and input vector have the minimum distance, in other words, the most active neurons. The algorithm will regulate the corresponding weight of the winners and form a new vector to make it closer to the input vector. Then the strongest region stimulated by the Kohonen learning rule is:

$$w_i(t+1)=w_i(t)+\alpha(x_i(t)-w_i(t)) \quad (16)$$

Here α is the constant for the study. It automatically adjusts, until the maximum principal component of the weight vector and input vector are coincident.

It is shown that the Kohonen learning rule is a cluster method without an instructor, which maps the formed centre of the cluster to the output plane, meanwhile maintainind an u nchanged topology. This is a two-layer network; the input of the neurons in the output layer is the dot metrix of the input vector *x*and weight vector *w*.

Boltzmann Rule

The Boltzmann rule is a strengthened learning rule, with a objective function to evaluate the output of the network. Compared with other learning rules, the Boltzmann learning rule is unique in that it uses a probability function to change the link weight.

It is known that the learning is to change the weight coefficients of neural networks. The energy function of the system is:

$$E=-\frac{1}{2}\sum_{i,j} w_{ij}\bullet x_i\bullet x_j \quad (17)$$

The state of energy E is determined by these weight coefficients w_{ij}, thereby changing the weight coefficients will lead to the changes of E. By-(5.17), it is known that the changes of Ewill directly lead to the changes of probability of the various states (especially at the minimum point of all the energy).

Competition Learning

In the learning, every neuron competes with each other to get the final winner node which has the maximized output. Only the winner node d_j^* and its neighbor nodes N_c can be adjusted, and its neighborhood gradually decreases in the learning. Supposed that there are n nodes in the input layer and m nodes in the output layer, the measurement of distance d_j can be denoted as:

$$d_j = \sum_{i=1}^{n} \left(x_i - w_{ij}\right)^2 \qquad i \in \left\{1,2,\cdots,n\right\}, j \in \left\{1,2,\cdots,m\right\} \tag{18}$$

$$d_j^* = \min\left(d_j\right) \qquad j \in \left\{1,2,\cdots,m\right\} \tag{19}$$

The learning rule of the winner node and its neighbor nodes N_c can be denoted as:

$$\Delta w_{ij} = \begin{cases} \alpha\left(x_i - w_{ij}\right) & j \in N_c \\ 0 & else \end{cases} \tag{20}$$

where α (α>0) is learning rate.

FUZZY SYSTEM

In 1965, Zadeh put forward the fuzzy set which became the principal method of dealing with various real-world objects. Since then, research and practical applications of the theory of fuzzy set and fuzzy signal processing have been widely carried out. Fuzzy Control and decision support systems are two prominent research and applications fields. This section will briefly introduce the basic concept of fuzzy math, algorithms, fuzzy logic and fuzzy sentences. These elements constitute the basic knowledge of fuzzy logic. Fuzzy calculation is based on fuzzy logic.

Fuzzy Set and the Membership Function

Suppose U is the collection of certain objects, called the domain, which is continuous or discrete: u is the element of U, denoted as U = (u) (Momoh, 1995).

Definition 1 (fuzzy sets) every mapping μ_F from domain U to interval [0,1], that is: U → [0,1], has identified a fuzzy subset F of U: called μ_F the membership function or grade of membership. That means μ_F is the extent of grade of F.

In the domain U, fuzzy subset is expressed as a dual sequence collection of the element u and the grade of membership μ_F, denoted as:

$$F=\{(u, \mu_F(u))|u \in U\} \tag{21}$$

If U is continuous, then fuzzy set F is denoted as:

$$F = \int_U \mu_F\left(u\right)/u \tag{22}$$

If U is discrete, the fuzzy set F is denoted as:

$$F = \mu_F\left(u_1\right)/u_1 + \mu_F\left(u_2\right)/u_2 + \cdots + \mu_F\left(u_n\right)/u_n = \sum_{i=1}^{n} \mu_F\left(u_i\right)/u_i, i = 1,2,\cdots,n \tag{23}$$

Definition 2 (fuzzy-set, the intersection and the fuzzy single point) if the fuzzy set is made up of all the element u which meet $\mu_F(u)\succ 0$, then call the set the fuzzy-set of fuzzy set F. When u meets $\mu_F(u)$=1,0, then call the fuzzy set the fuzzy single point.

Definition 3 (fuzzy relationship) if U, V are two non-empty fuzzy sets, then the first fuzzy subset R is called the fuzzy relationship from U to V, which is expressed as:

$$U \times V=\{((u,v), \mu_R(u,v))|u \in U, v \in V\} \tag{24}$$

Definition 4 (normal fuzzy sets, convex fuzzy sets and fuzzy number) fuzzy set F is in the real number domain R. if its membership function meets

$$\max_{x \in R} \mu_F\left(x\right) = 1$$

Then F is a normal fuzzy set; if any real number x (a<x<b) meets

$$\mu_F(x) \geq \min\{ \mu_F(a), \mu_F(b)\}$$

Then F is the convex fuzzy set; if F is normal and is convex, then F is a fuzzy number.

Definition 5 (language variant) variables are defined as a multi-language group $(x, T(x), U, G, M)$, where x is the variable name; $T(x)$ is the word set of x, that is the collection of the name of the language; U is the domain; G is the grammar rules generating the name of the language ; $M$is the grammar rules related to the meaning of the language value.

Each language value of the language variables is related to a fuzzy number of the domain U. The basic word set of the language variable is used to associate the fuzzy concept with the precise value, so as to quantify the qualitative concept and blur the quantitative data qualitatively (Mendel, 2007). For example, in an industrial furnace fuzzy control system, regard the temperature as a language variable, its word T (temperature) is recorded as T (temperature) = (ultra-high, high, high, medium, low, low, too low).

Fuzzy Logic and Fuzzy Deduction

Fuzzy logic deduction is based on fuzzy logic. It is an uncertain method of reasoning based on logical syllogism. On the presupposition of a fuzzy decision, it uses fuzzy language rules to get a similar judgment of the fuzzy conclusions. The fuzzy logic method is still in the research and development stage. There exist among others Zadeh law, Baldwin law, Tsukamoto law, Yager law and Mizumoto law. Zadeh law only is introduced here.

In fuzzy logic and approximate reasoning, there are two important rules of fuzzy reasoning: eneralized modus ponens (GMP) and generalized modus tollens (GMT), respectively called the general forward reasoning method and general backward reasoning method (Tanaka, 1993).

GMP reasoning rules is expressed as:

Premise 1: x is A'

Premise 2: if x is A, then y is B

Conclusion: y is B' (25)

GMT reasoning rules is expressed as:

Premise 1: y is B

Premise 2: if x is A, then y is B

Conclusion: x is A' (26)

In the above two formula, A, A', B and B' are fuzzy sets. x and $y$are language variables.

When $A = A'$ and $B = B'$, GMP will degenerate into positive premise deduction. It is closely related to positive data-driven reasoning, and is very useful in fuzzy logic control. When $A' = \bar{A}$ and $B' = \bar{B}$, GMT will degenerate into negative premise deduction. It is closely related to reverse goal-driven reasoning, and is very useful in the expert system.

Since Zadeh introduced multiple reasoning rules into approximate reasoning, dozens of implied functions with fuzzy variables have emerged, divided into three categories: fuzzy collection, fuzzy extraction and fuzzy implication. On the basis of the definition of the three, using a triangular paradigm and triangular co-triangular, the fuzzy implication relations generally used in fuzzy deduction are obtained.

Fuzzy Neural Networks and Fuzzy Neural Systems

Neural networks and **fuzzy systems** are two mainstream intelligence technologies (Sung, 2006). They can simulate human intelligence and do not require a precise mathematical model, can solve many complex, uncertain, non-linear problems that are impossible to cope with using traditional technology, and are easy to realize using hardware or software. With the further research of neural networks and fuzzy systems, the independent relationship of the two areas has gradually changed. The shortcomings of one area are compensated for by the other's advantages. Introducing fuzzy technology into neural networks widely enhances the information-processing coverage and capability of the neural networks, in processing precise

information and fuzzy information, and realizing precise association mapping and imprecise association mapping, which is fuzzy association mapping. Using a neural network to carry out fuzzy information processing may achieve the automatic extraction of fuzzy rules and the automatic optimization of the membership function, so the fuzzy system becomes an adaptive system. The integration of the neural networks and fuzzy systems, that is the integration of sign logical methods and mechanisms connected methods, and the combination of the numerical logic method and fuzzy logic method, have advantages over individual studies. The integrated system of neural networks and fuzzy systems not only has the ability of neural networks to learn, optimize and connect and structure, but also has the fuzzy systems' if-then rule, which is similar to the human way of thinking, and it is easy to embed the expert knowledge. The combination of the two is an inevitable development trend.

Neural Networks and Fuzzy Systems Integration

Integrating neural networks and fuzzy systems combine the advantages of both, so that neural network or fuzzy system expands its capabilities while maintaining the original function, not only by having the exact fitting and learning ability of a neural network but being also easy to understand by people, just as fuzzy model. The expert knowledge is switched into neural networks using a fuzzy model, and the neural network is then applied to a target system, in operation, through study, to improve the accuracy of the model. After learning, the parameters of the neural network switch into a fuzzy model to help explain the knowledge stored in the neural network(Kosko, 1992).

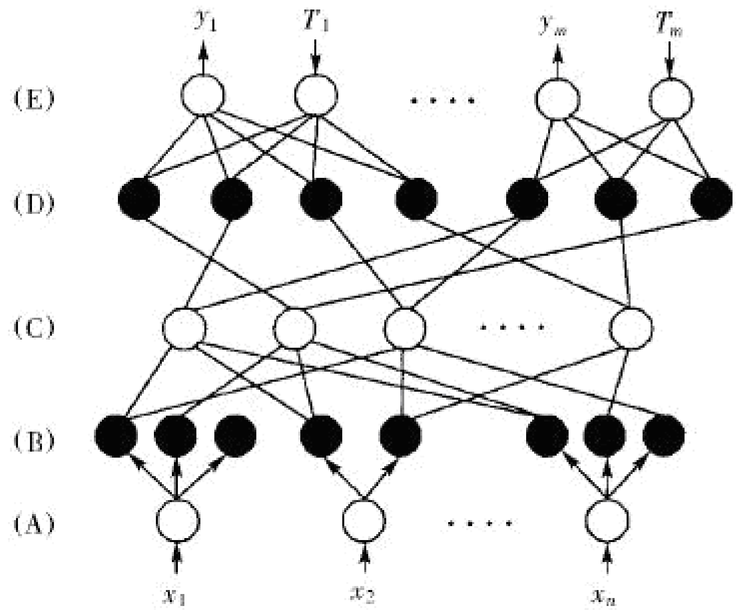
Initially, the target system model is based on fuzzy rules and is somewhat rough. After the learning of a cooperative system, the target system model is further improved. The system's structure and work process are as follows:

1. According to the expert knowledge of the target system, get the grade of membership and the fuzzy rules which can construct a fuzzy system;
2. According to the fuzzy system, establish the ways of connection and the connected weight of the neural network;
3. Apply the identified neural network to the actual target system;
4. Get the data from the target system, use the data to carry through the training and learning of the neural network to improve its accuracy;
5. Explain the changes of the connected weight of the neural network after learning, so as to enhance the understanding of the internal process of the neural network.

The integration of neural networks and fuzzy systems are generally divided into the following two ways:

1. **Fuzzy neural networks:** This is the combination of fuzzy system and neural network, which inserts fuzzy logic into the neural network so that the neural network has a logic reasoning functions, and improves the learning speed of the neural network using fuzzy logic. Fuzzy neural networks retain the basic nature and structure of neural networks, except that some of the components will be 'vague', but their nature is that of neural networks, so they are mainly used in the field of pattern recognition.
2. **Neuron-fuzzy system:** This is the combination of neural network and fuzzy system; the learning capability of the neural network is introduced into the fuzzy system so that the fuzzy system can automatically get fuzzy rules from learning. In neuron-fuzzy systems, neural networks process the expanded number of the fuzzy set, for example, the selection of the membership functions and the realization of mapping between fuzzy

Figure 2. Neural network model based on fuzzy reasoning



sets (for the fuzzy rules). Because the nature of the fuzzy system is a fuzzy logic system, it is mainly used in control areas.

3. **Attempts to integrate neural networks and fuzzy systems are mainly in the following areas:** Fuzzy clustering networks, neural networks based on fuzzy reasoning, adaptive *fuzzy control* systems, adaptive fuzzy associative memory systems and fuzzy systems modeling based on neural networks.

Fuzzy Neural Network (FNN)

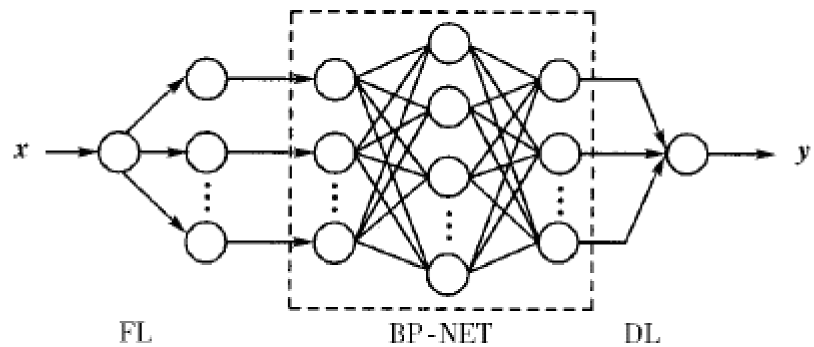
Introducing fuzzy logic technology into a neural network is known as a Fuzzy Neural Network (FNN). In fact, FNN is a blurred neural network, which means that the components of the neural network (input, output, transfer function, weights, learning algorithm, etc.) are blurred (Dutra, 2008).

A neural network model based on fuzzy deduction is given in Figure 2. X is the input, Y is the output, T is the signal from the teacher, the white nodes are linear nodes, and the black nodes are the Sigmoid nodes. The system is divided into five levels: A layer is the input layer, directly transmitting the input to the next level. B layer is the input language variable layer, the output of which is membership functions. C layer is the rule layer, realizing the match of the prerequisite of fuzzy rules, and implementing the fuzzy 'and' operation. D layer is the output language variable layer, implementing the fuzzy summation operation. E layer is the judgment layer, generating output. Guided study and training methods are used in such fuzzy neural networks.

Fuzzy System Modeling Based on Neural Network

Fuzzy system modeling (Fuzzy Modeling) is an approach using fuzzy reasoning to describe complex characteristics of nonlinear systems. Fuzzy modeling based on the neural network system introduces the technologies of the fuzzy neural network into the fuzzy expert system, the core of

Figure 3. Fuzzy system modeling based on BP neural network



which is to use the neural network fuzzy system to achieve the input / output mapping. Its general structure and function are similar to the general neural network expert system, but the difference is that it stores fuzzy model or models. In the practical fuzzy expert system applying neural network technology, it is therefore necessary to consider the following key question: how to achieve the fuzzy model or the model of fuzzy associative memory storage, and how to achieve the associative memory of fuzzy model or models and how to achieve the fuzzy association and fuzzy mapping of the fuzzy model or models.

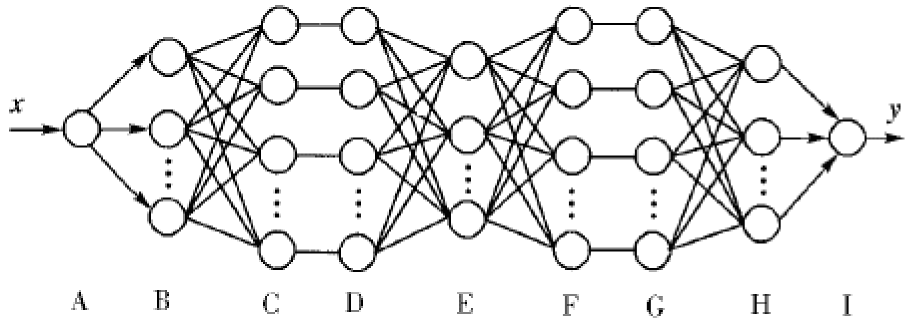
In order to achieve the storage and mapping of the fuzzy model, which is to achieve the conventional sense of storage and use of fuzzy rules, a feed-forward neural network could be introduced into fuzzy reasoning, optioning a feed-forward neural network to construct a fuzzy deduction system, generally using the fuzzy point-point mapping model. The basic idea is to apply fuzzy division to the input variables space, so as to form a point space made up of finite fuzzy points. This process is known as a 'fuzzy' process, whose principles are as follows:

Suppose X is the input variables space of the network, X will be divided into a number of fuzzy spaces $X_i(i=1,2,...,N)$ known as: 'vague points'. After fuzzy division, variable space X transforms into a fuzzy point space composed of finite vague points, and any object in X is represented of these vague points. The fuzzy deduction system is characterized by $CY(y)=G(CX(x))$ and $CY(y)$ and $CX(x)$ respectively represent the corresponding fuzzy point vector of the output and input of the system. G is the fuzzy mapping from x to y. after inputting x, the output vector is got through $CY(y)=G(CX(x))$ and further using 'space center' to make the result non-fuzzy.

Taking the BP network as an example, this type of fuzzy neural network structure is shown in Figure 3, in which FL is the fuzzy layer, BP-NET is the feed-forward type network layer, DL is the non-fuzzy layer, each node of the FL layer corresponds to the fuzzy point of input variables space, the output of FL is the fuzzy point vector corresponding to input variables, and the output of the feed-forward neural network is the grade of membership of y corresponding to a fuzzy point in its space Y. FL transforms the input of the system into a fuzzy point vector, the feed-forward neural network realizes the fuzzy mapping, the DL layer transforms the output fuzzy point vector of the feed-forward neural network into the determinate output of the system. As to such fuzzy neural networks, the neural networks using fuzzy mapping require pre-training with the patterns of samples, rather than training with an entire network.

Neural networks are further introduced to fuzzication, fuzzy reasoning, fuzzy judgment modules, which use three neural networks to simulate the input and output of the three modules.

Figure 4. Improved fuzzy system modeling based on BP neural network system



The structure of such a fuzzy neural network is shown in Figure 4, in which A, B, C layers realize fuzzy, D, E, F layers realize fuzzy reasoning, G, H, L layers realize non-fuzzy, the number of C, D-nodes are determined by the number of fuzzy sets of the input variables, the number of F, G nodes are determined by the number of fuzzy sets of the output variables. Such FNN have higher accuracy than the former FNN.

EVOLUTIONARY COMPUTATION

Basic Elements and Mathematical Description of the Evolution System

In this section, we will give the formal model of the simulated evolutionary algorithm (SEA), and introduce the mathematical description of the basic element involved in SEC, such as the choice operator, hybrid operator, variation operator and such.

Choice Operator

Definition 1 suppose the parameters $0<\alpha<+\infty$, choice operator is a random map selecting an individual from a group, recorded as

$$T_s^\alpha : S^N \to S$$

In particular, in accordance with the rules of probability

$$P\left\{T_s^\alpha\left(A\right) = A_t\right\} = \frac{f^a\left(A_t\right)}{\sum_{j=1}^{N} f^a\left(A_j\right)} \quad (27)$$

The choice operator T_s^α in the definition is known as the α-order fitness choices operator; When α=1, T_s is known as the proportional choice operator.

Definition 2 symmetrical product space

$$S^2=\{(A_1,A_2)|A_t\in S, i=1,2\}$$

It is the mother space, the elements (the individual) (A_1,A_2) in the mother space is called the mother; the random map from symmetrical group space S^N to the mother space S^2

$$\tilde{T}_s^a : S^N \to S^2$$

It is the choice operator of the mother; particular, in accordance with the rules of probability

$$p\left\{\left(A_i, A_j\right)\right\} = P\left\{\tilde{T}_s^a\left(A\right) = \left(A_i, A_j\right)\right\} = \frac{f^\alpha\left(A_i\right)}{\sum_{k=1}^{N} f^\alpha\left(A_k\right)} \bullet \frac{f^\alpha\left(A_j\right)}{\sum_{k=1}^{N} f^\alpha\left(A_k\right)} \quad (28)$$

The maternal choice is known as the αOrder choice operator of the mother.

Hybrid Operator

Definition 3: a map from the mother-space to the individual-space

$T_c: S^2 \rightarrow S$

It is a hybrid operator. In particular,

1. If identifying a gene as a hybrid randomly, exchange two individuals of the mother in the second-half from hybrid points, then have two new individuals, from which choose the first one as the result of hybridization, T_c is called a single point hybrid operator.
2. If exchanging the latter part of two hybrids in all the single point hybridization, with the probability p_c, then T_c is called a single point random hybrid operator.
3. If exchanging the corresponding component of the first individual and the second one in the mother independently, with a hybrid probability p_c then T_c is called uniform random hybrid operator.

Variation Operator

Definition 4: a random map from an individual-space to another one

$T_m: S \rightarrow S$

It is called a variation operator, if T_m changes the value of the individual independently with the variation rate p_m.

In a genetic algorithm design, generally, hybrid probability has a greater value between 0.6 and 0.9; variation probability has a smaller value between 0.001 and 0.01.

Delete Operator

Definition 5: a real number $0 \leq \alpha < \infty$, a random map that removes a individual from a group

$T_d^{\alpha} : S^N \rightarrow S^{N-1}$

It is the removal operator for the αorder, if T_s^{α} has the probability distribution

$$P\left\{T_d^{\alpha}\left(A\right) = A \setminus \left\{A_i\right\}\right\} = \frac{f^{-\alpha}\left(A_i\right)}{\sum_{j=1}^{N} f^{-\alpha}\left(A_j\right)} \quad (29)$$

$A=(A_1,A_2,\ldots,A_N)\in S^N$, especially when α=1, T_d^1 is known as the proportion omissions operator; When α=0, then T_d^0 is known as the Uniform addition operator.

Delete operator is designed to reflect the nature of 'the inferior', and the choice operator is designed to reflect the nature of 'winning'. The following theorem will note that when αis great, the greater their fitness is, the easier the individual to be retained. When αsmall, the relationship between the delete operator and Vietnam is not obvious but random.

Genetic Algorithm

Simple Principles of the Genetic Algorithm

The genetic algorithm (GA) is an efficient exploration algorithm based on the natural group genetic evolution mechanism. It abandons traditional search methods, simulates the natural process of biological evolution, and makes use of artificial evolution to search on the target space randomly. It regards the probable result in the problem domain as an individual or group of chromosomes, and encodes them into a string of symbols, simulates Darwin's genetic selection and natural selection process of biological evolution, uses genetic op-

erations (genetic, crossover and variation) on the group repeatedly, and evaluates each individual in accordance with the fitness function of the objectives. Based on the survival of the fittest, the evolution of the survival of the fittest rules; better groups will occur constantly, at the same time, and a search for the best individuals in groups by the overall parallel ways is carried out to seek the required results(Werner, 2007).

Let us illustrate the principles of the genetic algorithm firstly through an example:

Assume that the greatest value of the function $f(x)=x^2$ is required to gain; x is a natural number, $0 \leq x \leq 31$. Now, regard each as alive, through evolution, capable of surviving in the end.

Coding

Each number is regarded as alive, and each is given certain genes; this process is called encoding. The variable x is encoded into a 5 long unsigned binary integer, such as x = 13 is expressed as 01,101, that is, the genes of 13 is 01,101.

Formation of the Initial Group

As genetic needs, set some initial biological groups as the first generation of biological reproduction. It should be noted that each individual of the initial group is generated randomly, which will ensure the diversity and fair competition of biology.

Fitness Assessment Test

The evolution of biology subjects the gene to the 'survival of the fittest' therefore, the measure of which genes are 'excellent' and which are 'bad', is called 'fitness'. Clearly, as the maximum of $f(x)=x^2$ is requested, therefore, the gene that makes the function have the maximum value is excellent, otherwise the gene is bad, so $f(x)=x^2$is defined as the fitness function, used to measure a certain ability of organisms to adapt.

Choice

Next, carry out the fittest process of the survival, a process in the genetic algorithm, called the choice. Note that choice should be a random process; poor genes may not necessarily be eliminated, but the probability of being eliminated is relatively high, which is the same within the laws of nature. Therefore, choice takes the form of gambling.

Cross-Operation

Next carry out the cross-breeding by randomly selecting two organisms; let them exchange part of their genes, which will create two new organisms, for the second generation.

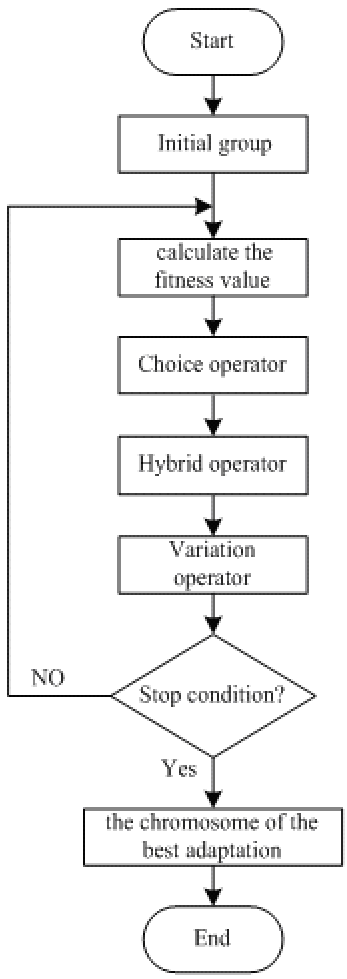
Variation

In the biological situation, there are also genetic variations. Here, variations are introduced, so that the genes in the organism change with a certain degree of probability. Such introduction of appropriate disturbance can avoid local extremes. The simplest algorithm of the above method is the genetic algorithm. Through the above steps organisms are evolved constantly, and the genes of the organisms gradually tend towards the optimal, finally achieving the desired result. A simple genetic algorithm diagram is shown in Figure 5.

The simplest stop conditions of the algorithm are the following two: 1) if completing the given algebra of the evolution, then stop; 2) if several generations of optimal individuals in a row have no improvement in average fitness, then stop. The major steps of the general genetic algorithm are as follows:

1. Randomly generate a initial composition of the group made up of a characteristic string of a certain length;
2. Repeat the following step i) and ii) for the string group, until the criteria of the stop are met:
    a. Calculate the fitness of each individual string of fitness;
    b. Generate the next generation group using the copy operator, variation operator and cross.

Figure 5. Diagram of a simple genetic algorithm



3. Designate the best individual string to appear in future generations as the results of the implementation of the genetic algorithm.

The Characteristics of the Genetic Algorithm

The genetic algorithm is a space-based search algorithm, which through natural selection and genetic variation, such as operations and Darwin's theory of the survival of the fittest, simulates the natural evolutionary process to seek the answer to a question. The solving process of the genetic algorithm is seen as an optimization process. It must be pointed out that a genetic algorithm can not guarantee that all the results are the best answers, but in certain ways, error is controlled within limits. Genetic algorithms have the following characteristics:

1. a genetic algorithm is the coding of the parameters set rather than the evolution of the organism;
2. a genetic algorithm begins from the code group rather than searching from a single solution;
3. a genetic algorithms uses the adaptability of the target function rather than the derivation or other supporting information to guide the search;
4. genetic algorithms use selection, crossover, variation, and other operators rather than using certain rules to carry out a random operation.

The genetic algorithms uses simple coding technology and a reproduction mechanism to demonstrate complex phenomena to solve very difficult problems. It is free of the restrictive assumptions of the search space, such as not necessarily requiring the continuity, or existence of derivative and a single peak assumptions. It can find the optimal solution of the whole with high probability. Because of the inherent parallelism, the genetic algorithm is applicable to large-scaled parallel computing and is widely applied to optimal, machine learning, parallel processing fields, and other fields.

Typical Implementation Strategy of Evolutionary Computation

The Algorithm Model of Evolution Strategy

The simplest form of evolutionary strategy is described as follows:

1. Problem is defined as seeking the real n-dimensional vector x associated with the extremes of the function: $F(x):R^n \rightarrow R$.
2. Select the initial group of the parent vector from every possible range. The distribution of the initial test is of typical consistency.
3. Parent vector $x^i, i=1,2,...,p$ generates the sub-vector by adding a Gaussian random variable of zero-sum standard deviation and pre-selecting the standard deviation of x.
4. Select and determine which vectors should be maintained by sorting the error $F\left(x_i'\right)\left(i=1,2,\cdots,p\right)$. Those vectors P with the smallest error are the new parent of the next generation.
5. The process of generating the new test data and selecting the smallest error will continue until the answer to the requirements is found or all the calculations are completed.

In this model, the component of the test answer is seen as the behavior of the individual, rather than the gene of the chromosomes. The genetic origin of mixed characteristics is hypothetical, but their liaison mechanism is not described in detail. This evolutionary strategy assumes that, no matter what genetic changes emerge, each of the changes brought about by the behaviors will distribute in accordance with the Gaussian distribution with zero standard deviation and some standard deviation distribution(Dumitrescu, 2000). With more efficiency and diversity, certain genetic changes can affect many mixed characteristics, thus while creating a new sub-group, the composition of the parent is also changed.

At first, an evolutionary strategy should detect the aforementioned method, but focus on a single parent-single offspring search. Such a single parent-single offspring search is represented by (1 +1) - ES, in which the single offspring is produced by the single parent. Both of them compete with each other to survive the survival of the fittest rule. This method used to optimization algorithms has two defects: 1) the standard deviation (average step) of every dimension slows the speed of convergence to the optimal value; 2) the instability of the point-to-point search may cause stagnation in the local minimum value.

The Difference Between Evolution Strategy and the Genetic Algorithm

Through the study of evolutionary strategy, it is shown that evolutionary strategy and genetic algorithms have strong similarities, as a kind of algorithm imitating the evolutionary principles of nature. At the same time, there are also differences between them, the main one being in their different areas of research. Evolutionary strategy is a numerical optimization method, which adapts the method of mountaineering with adaptive steps and particular inclination. Until recently, evolutionary strategy has been applied to discrete optimization problems. The genetic algorithm is an adaptive search technology in the broad sense. It decides how to allocate test data which is of exponential growth above the average plan. Genetic algorithms have been used in many fields; parameters optimization is only one of its applications.

In addition to the difference in the research and application, there are some other differences between the evolutionary strategy and genetic algorithms.

1. Their ways of representing the individual are different. Evolutionary strategy runs in the floating-point vector, while the genetic algorithm runs in the general operation of the binary vector.
2. The selection process of evolutionary strategy and genetic algorithm are different.
3. The replication parameters of evolutionary strategy and genetic algorithm are different; the replication of genetic algorithm parameters (the possibility of crossover and variation) in the process of evolution

is constant, while evolution strategy needs to constantly change them.

With the development of technology, the difference between genetic algorithms and evolutionary strategy will become less and less.

CASE STUDY

In this section, we will present a case study about integrated computation intelligence method for complex multi-objective optimization manufacturing process. Artificial Neural Network was used for system modeling, which searched fitness function and target function value for GA. And GA was applied in multi-objective optimization.

Background and Motivation

An ideal mechanical manufacturing procedure is characterized by using the lowest cost, the smallest rejection ratio, the least tool abrasion, and the smallest resource consumption to achieve the highest production, the best manufacture accuracy, and the best surface texture. But due to the complexity of the processing, the non-linear of the system and the alteration of the process parameter and so on, it's difficult for the traditional numerical calculation and the experimental fitness method to achieve the multi-objective optimization decision in the processing. So it's significant to establish an effective system framework and explore some new algorithms to solve problems emerging under complicated manufacturing condition.

In the past, most multi-objective optimization schemes were based on mathematic models. Because of the difficulty of mathematic modeling, and the difference between the simplified and assumed modeling condition and the real world, the optimization effectiveness is greatly decreasing. Although many researchers have tried to apply ANN to deal with the multi-objective optimization problem, most only built single frameworks with ANN, but not hybrid systems. So the black-box deficiency of ANN still unresolved. Recently, many researchers try Genetic Algorithm (GA) to deal with the optimization. The toughest task with this method is to find out a fitness function to evaluate the output of the optimization. In general, the fitness function is defined using mathematic models or probability distribution function, but it's never easy to find a proper fitness function by traditional method, which leads to the limitation to the widely application of GA in the manufacture system.

Therefore some researchers turn to the hybrid system named integrated intelligent system or mixed intelligent system (WANG, 2003). The following are the recent popular mix and integration techniques. One is the integration of fuzzy logic and ANN, which means that the system is supported by the fuzzy logic and the fuzzy logic rule is implemented by ANN, or the structure of ANN is adopted in the system and the fuzzy logic is used in the reasoning process; one is the integration of GA and ANN, which optimize ANN using GA; and another is the integration of GA and fuzzy logic, which improve the GA performance using the regular membership function of fuzzy logic. A model of complicated making process is proposed applying integrated computational intelligence and the multi-objective optimization problem is resolved from an entirely new perspective to achieve the simplification, accuracy and effectiveness of the system.

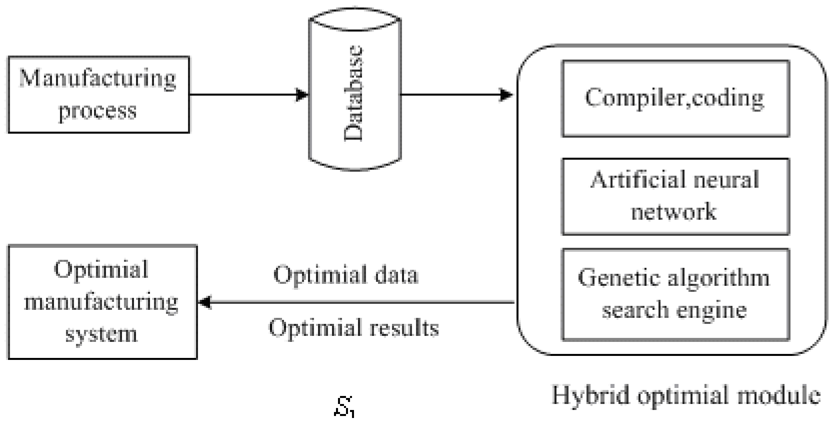
Modeling

The Research Scheme of the Multi-Objective Optimization System

Integrated computational intelligence technology is used to optimize the multi-objective system, the research scheme is showed in Figure 6.

1. **Data acquisition and processing:** using sensor to collect the data in the manufacture,

Figure 6. Integrated computational intelligence based multi-objective optimization system of manufacturing process



and processing and transforming the noise-mixed data properly, then storing the noise-free data into the database.

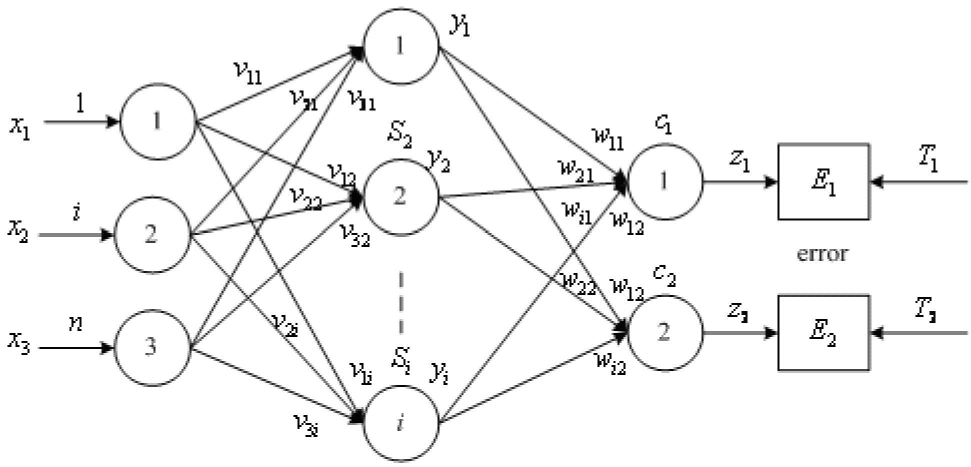
2. **ANN:** because most of the past optimization models were based on approximate functions, it's difficult to get a continuous function between the input and the objective in manufacturing process. ANN is superb in the modeling. It can use much less practically measurable data to establish a complete and effective model. So ANN can be used to deal with the system modeling. When the system model is based on ANN, the output of the system must be representative of the manufacturing performance, so as to build up the model which carries out the mapping from multi-input to the output of the manufacturing performance.
3. **Hybrid optimization module:** traditional optimization methods usually solve the single-objective optimization problem which meets the continuous precondition. The computational capabilities are increasing exponentially with the development of the computer technology, which gains GA a lot of attention.

A hybrid intelligent system can use the advantages of the intelligent technology and avoid its disadvantages, so as to improve the calculation efficiency of the system. A model of the input and the objective is proposed to serve as fitness function to do optimization by integrating GA and ANN, so there is no need to form a new fitness function.

Modeling of the Optimization Structure of Multi-Objective Parameters

ANN and GA are integrated in this case. The details are encoding the Back Propagation (BP) ANN, doing the learning process using GA, dynamically adjusting the connection weights and thresholds, randomly generating the initial species group, carrying out intersection and variation until the training of the network is finished. The random selection of the initial species group can make sure every state is traversed, so as to ensure the survival of the optimum solution in the evolution. But it is the random selection that enlarges the data of the evolution and belongs the time of evolution. In our research, GA is mainly used to optimize the result of the training of ANN, and the optimized result is used to evaluate the system.

Figure 7. The structure of neural network based on GA



Multi-objective problem can be depicted as follows:

$T_{\text{-min}} f(x)=[f_1(x), f_2(x),...,f_p(x)]^T,\ x \in X, X \subseteq R^m$

where $T_{\text{-min}}$ is vector minimization, this is, each sub-objective function in vector objective $[f_1(x), f_2(x),...,f_p(x)]^T$ should be the smallest as far as possible.

In practical programs, the objective and the parameter are correlated and countered. The relationship function of the objective and the input is difficult to precisely express by the mathematic method. Additionally, there is conflict among the multiple objectives. The improvement of one sub-objective may result in damage to another one's performance, which means that it's impossible for all sub-objectives to achieve the optimum value. What we can do is to coordinate them to achieve a compromise in which each sub-objective does its best to approach the optimum. The processing quality takes the priority over the time and benefits in the objective balance. Accordingly, GA is what is exactly needed to solve the problem.

An analytical example is showed in Figure 7. The network contains 3 layers. The input layer contains 3 nodes, the output 2, and the hidden layer unknown and optional. The input vector and the output vector are respectively denoted as:

$X=[x_1, x_2, x_3]^T$

$Z=[z_1, z_2, z_3]^T$

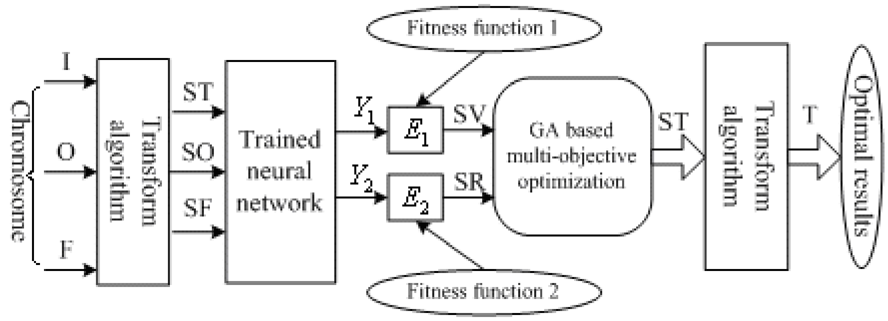
The steps of training are as follows:

Step 1 initialize. The connection weight and threshold of each node is assumed to be random. The random initial species group is $Q=\{q_1, q_2,...,q_n\}$ the number of which is n. Here q_i is made up of a set of weight vector and threshold vector, namely $q_i=\{V_i, W_i, S_i, C_i\}$.

Step 2 choose an encode method. An encode method is defined to show the scheme of ANN. The general encode methods are using binary code. After chromosomes being encoded, n individuals are generated to constitute the initial species group.

Step 3 calculate the adaptability. According to the random weight and threshold, the real output of ANN is calculated through the structure of the network, namely $Y=VX\text{-}S_i$ and $Z=WY\text{-}C_i$, the sum of squares of the difference between the real output Z and the expected output T, $E_{se} = \sum_{i=1}^{n} \sum_{1}^{3} \left(T - Z\right)^2$

Figure 8. The structure of multi-objective parameters optimization based on neural network and GA



is called the global error which is regarded as the adaptability of GA and used to judge the quality of individual.

Step 4 optimization calculations. Design the choice policy of GA, set the parameters of GA, the process of calculation is as Figure 8, finally acquire the acceptable individual q_m.

Step 5 optimize network structure. The number of the hidden layers is not certain in this model. In the training, it has great effect on the training process and result, and it can also be optimized. First, set a certain evaluation criteria, for example, using the calculation time within which the global error diminishes to the acceptable range, to be the evaluation criteria, then compare the calculation time of different layers to choose the least one to be the number of the hidden layers of the network .

Step 6 decode. Decode q_m to get a group of parameter Q_m of the trained ANN. The trained ANN is composed of $Q_{m.}$

As set forth, the biggest problem with GA is to find out a fitness function to evaluate the result of the optimization. Most researchers apply mathematic models or probability distribution function to define the fitness function and using it to evaluate the output. However, in this case, we directly use ANN module to be the fitness function, so there is no need to form a new fitness function, which can effectively solve the above problem and elevate the efficiency of calculation.

Therefore, compared with traditional methods, the proposed one has following advantages:

1. It can be widely used in almost all optimization algorithms, and needn't establish any traditional mathematic relationship. Although, it doesn't as accurate as traditional method in some respects, it's enough to get a required solution.
2. The complex relationship of the parameters can be determined. For example, in mechanical workout, though the relationship of the parameters can be found out by the analysis of experimental data, it's just a kind of qualitative analysis. However using the proposed method in this case, it is easy to get the proper quantitative analysis of the relationship of the parameters.

Conclusion

The integrated model has the advantages which are not for other kinds of intelligent calculation method. 1) the advantages of ANN is fully used to resolve the difficulty of modeling of the complex systems. 2) bringing in GA effectively resolve the difficulty of learning of ANN and multi-objective optimization in the complex systems. Verified by experiment, the optimized result can not only better control the surface roughness, but also improve the productivity of the machine.

SUMMARY

In This chapter, we firstly introduced computational intelligence, with reference to neural computation, fuzzy, evolutionary computation as the main research field of computational intelligence.

Artificial neural network has advantages of parallel distributed processing, nonlinear mapping, learning through learning, adaptation and integration, and the easy realization of hardware, parallel processing and so on. The basic element of the neural network is the neurons, with a number of inputs and outputs. Between the neurons are the connections with the weight value. The input signal transforms into output by the stimulating function. Artificial neural network is divided into two basic network structures - recursive (feedback) network and multi-storey (feed forward). An artificial neural network is divided into supervised study and non-supervised study in terms of learning algorithms. In fact, it is seen as a special case of supervised study. There are many artificial neural network models, the most widely used models are BP networks and Hopfield networks. The artificial neural network is used for knowledge representation and reasoning, and has gained extensive application.

Definitions of fuzzy set, fuzzy logic and fuzzy deduction were given in this chapter, then a fuzzy neural network structure was studied combining artificial neural networks with fuzzy logic principle. Neural computations based on neural network and fuzzy computations based on fuzzy logic are all based on numerical calculation, which are the important components of computational intelligence.

In evolutionary computation, the genetic algorithm and evolutionary strategy were studied, in which the survival of the fittest rules of nature apply, imitate the evolutionary mechanisms of organisms, and are applied to the optimization problem of complex system. The genetic algorithms is a kind of artificial search algorithm imitating the biological genetics and natural selection mechanism, and is a mathematical simulation of the process of biological evolution, also the most important form of evolutionary computation. In this section, the structure and mechanism of a simple genetic algorithm were given, including the encoding and decoding of the algorithm, the fitness function, genetic operation and so on. In the discussion of the steps of the genetic algorithm, the characteristics of were summed up and the algorithm diagram and general structure of genetic algorithm were presented. Evolutionary strategy is an algorithm imitating evolutionary theory to solve optimization problems. A simple evolutionary strategy was first described, illustrated, and then the differences between evolution strategy and genetic algorithms were discussed.

As an emerging technology, computational manufacturing offers many promising applications in various research fields. However, its fundamental theories and methods need to be further investigated. Many problems with its supporting theories exist, such as computational geometry, Internet-based design, control and maintenance, computational intelligence, and virtual prototyping. In fact, these theories themselves are currently active research topics. The main purpose of computational manufacturing is to integrate all kinds of manufacturing information by establishing efficient computational models and methods for various computational problems in manufacturing systems and processes. This will be helpful in improving manufacturing companies' creativity and competitiveness.

REFERENCES

Agogino, A., Stanley, K., & Miikkulainen, R. (2000). Online Interactive Neuro-evolution. *Neural Processing Letters*, *11*, 29–37. doi:10.1023/A:1009615730125

Antonio, L. S., Amerieo, L. A., & Jorge, P. (2000). Distributed Planning and control systems for the virtual enterprise: organizational requirements and development life-cycle. *Journal of Intelligent Manufacturing*, *11*, 253–270. doi:10.1023/A:1008967209167

Cacciola, M., Calcagno, S., Morabito, F. C., & Versaci, M. (2008). Computational Intelligence Aspects for Defect Classification in Aeronautic Composites by Using Ultrasonic Pulses. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *55*(4), 870–878. doi:10.1109/TUFFC.2008.722

Cantu-Paz, E. (2000). *Efficient and Accurate Parallel Genetic Algorithms*. Amsterdam: Kluwer Academic Publishers.

Celano, G., Costa, A., & Fichera, S. (2000). Fuzzy scheduling of a flexible assembly line through an evolutionary algorithm. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, (pp. 328-333).

Chen, D., & Chen, Y. Z. (2004). Switched system optimal control based on dynamical programming principle and evolutional computation. *Intelligent Control and Automation, WCICA 2004, Fifth World Congress on*, (Vol. 2, pp. 1104 – 1108).

Christine, S. (2002). Ants can solve constrain satisfaction problems. *IEEE Transactions on Evolutionary Computation*, *6*(4), 347–357. doi:10.1109/TEVC.2002.802449

Clec, M., & Kennedy, J. (2002). The particle swarm explosion stability and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, *6*(1), 58–73. doi:10.1109/4235.985692

Clerc, M. (2000). *Discrete particle swarm optimization illustrated by the traveling salesman problem*. Retrieved from http://www. maurice-clerc.net

Dong, J., & Yang, G.-H. (2008). Dynamic Output Feedback Control Synthesis for Continuous-Time T–S Fuzzy Systems via a Switched Fuzzy Control Scheme. *IEEE Transactions on Systems, Man, and Cybernetics*. *Part B*, *38*(4), 1166–1175.

Dumitrescu, D., Lazzerini, B., & Jain, L. C. (2000). *Evolutionary Computation*. Boca Raton: CRC Press.

Dutra, R. G., & Martucci, M. (2008). Adaptive Fuzzy Neural Tree Network. [Revista IEEE America Latina]. *Latin America Transactions, IEEE*, *6*(5), 453–460. doi:10.1109/TLA.2008.4839115

Fogel, D. B. (2001). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence* (2nd ed.). New York: Wiley-IEEE Press.

Katayama, M., & Yamamoto, T. (2004). Evolutional computation in designing a robust IMC-based PID controller. *IEEE International Conference on Systems, Man and Cybernetics*, (Vol. 7) 6120-6124.

Khoo, L. P., Lee, S. G., & Yin, X. F. (2000). Prototype genetic algorithm-enhanced multi-objective scheduler for manufacturing systems. *International Journal of Advanced Manufacturing Technology*, *6*(2), 131–138. doi:10.1007/s001700050018

Kosko, B. (1992). *Neural Networks and Fuzzy System*. Upper Saddle River, NJ: Prentice Hall International Inc.

Liu, J., & Cai, Z. (2002). An incremental time-delay neural network for dynamical recurrent associative memory. *High Technology Letters*, *8*(1), 72–75.

Liu, J., & Cai, Z. (2002). Learning of goal directed spatial knowledge from temporal experience for navigation. In *Proc 6th Int Confon Intelligent Engineering Systems*, (pp. 57-62).

Mak, K. L., & Wong, Y. S. (2000). Genetic design of cellular manufacturing systems. *Human Factors and Ergonomics in Manufacturing*, *10*(2), 177–192. doi:10.1002/(SICI)1520-6564(200021)10:2<177::AID-HFM4>3.0.CO;2-A

Mendel, J. M. (2007). Type-2 fuzzy sets and systems: an overview. *Computational Intelligence Magazine*, *2*(1), 20–29. doi:10.1109/MCI.2007.380672

Ming, X. G., & Mak, K. L. (2000). Intelligent setup planning in manufacturing by neural networks based approach. *Journal of Intelligent Manufacturing*, (11): 311–331. doi:10.1023/A:1008975426914

Momoh, J. A., Ma, X. W., & Tomsovic, K. (1995). Overview and literature survey of fuzzy set theory in power systems. *IEEE Transactions on Power Systems*, *10*(3), 1676–1690. doi:10.1109/59.466473

Ritter, G. X. (2007). Towards a Unified Modeling and Knowledge-Representation Based on Lattice Theory: Computational Intelligence and Soft Computing Applications (Studies in Computational Intelligence). *IEEE Transactions on Neural Networks*, *18*(3), 965–966. doi:10.1109/TNN.2007.897476

Robinson, C. J., & Fogel, D. B. (2003). Computational Intelligence, (1st Ed.).

Shambu, G., & Suresh, N. C. (2000). Performance of hybrid cellular manufacturing systems: a computer simulation investigation. *European Journal of Operational Research*, *120*(2), 436–458. doi:10.1016/S0377-2217(98)00371-3

Srivastava, S.K., & Cartes, D.A., & Maturana. (2008). Control System Test Bed for Demonstration of Distributed Computational Intelligence Applied to Reconfiguring Heterogeneous Systems. *Instrumentation & Measurement Magazine*, *11*(1), 30–37. doi:10.1109/IM-M.2008.4449012

Sung-Kwun, O., Pedrycz, W., & Park, B.-J. (2006). Multilayer hybrid fuzzy neural networks: synthesis via technologies of advanced computational intelligence. *IEEE Transactions on Circuits and Systems. I, Regular Papers*, *53*(3), 688–703. doi:10.1109/TCSI.2005.857774

Tanaka, Y. (1993). An overview of fuzzy logic. In *WESCON/'93, Conference Record, Sept. 28-30 1993,* (pp. 446 - 450).

Wang, K., & Gelgelee, H. L. (2003). A hybrid intelligent method for modeling the EDM process. *International Journal of Machine Tools & Manufacture*, *43*(10), 995–999. doi:10.1016/S0890-6955(03)00102-0

Werner, D. H., & Haupt, R. L. (2007). Genetic Algorithms in Electromagnetics, (1st Ed., pp. 301).

Xu, Z. B., Leung, K. S., & Liang, Y. (2003). Efficiency speed up strategies for evolutionary computation: fundamentals and fast GAs. In *Applied Math* (pp. 341–388). Comput.

Chapter 6
Business Process Modeling and Information Systems Modeling

INTRODUCTION

Background and Motivation

Today's world is characterized by globalization and the rapid advance of information technology. To face such unprecedented change and survive, enterprises have to continuously review their products, services, and relations with the environment. Information systems, which have now become an integral part of business, are relied on to assess the quality of products and effectiveness of services.

Unfortunately, very often software system does not properly support businesses. The causes may be poorly-defined assessment of requirements, deficiencies in proper business understanding by the software design team, or even the nature of the business (which may change so often that the software simply cannot keep pace). According to Davenport and Short (1990), the relationship between business process design and information systems has never been fully exploited in practice.

DOI: 10.4018/978-1-60566-864-2.ch006

Modeling is an essential means to address the 'business/IS fit' problem. Within the field of organizational design and information systems development, approaches to requirements engineering involve a detailed modeling of different aspects such as system structure, data or behavior. Models enable decision makers to filter out the irrelevant complexities of the real world, so that efforts can be directed toward the most important parts of the system under study. Using a model, those responsible for a software development project's success can assure themselves that business functionality is complete and correct, end-user needs are met, and program design supports requirements for scalability, robustness, security, extendibility, and other characteristics, before implementation in code renders changes difficult and expensive to make. Furthermore, these models are an essential means of communication between system developers and expert users, and they are the basis from which system design and implementation are derived in later stages of the development process.

In order to support system development in an optimal way, description techniques for models of specific system views must be intuitively under-

standable in both business process design and information technology. On the one hand, the choice of a particular way of conducting business in an organization will influence the design and structure of the information systems to support this process. On the other, advances in information technology can generate completely new opportunities for organizations and thus influence the design of specific business process layouts (Giaglis, 2001).

Given such recursive relationships between business process and information technology, it is a decisive factor for the success of system development to align the design of information systems with the design of corresponding business processes. Successfully integrating these systems into the enterprise often requires modeling even the manual organizational processes with which these systems interact. Therefore, more and more people, both from the field of information technology and business engineering, have concluded that successful systems start with an understanding of the business process of an organization. To align the design of information systems with the design of the corresponding business process, conceptual models for both information and business process are essential. Conceptual modeling of business process is deployed to facilitate the development of information systems that supports business process, and to permit the analysis and re-engineering or improvement of them (Aguilar-Saven, 2004).

The increasing popularity of business process orientation has yield a rapidly growing number of methodologies, and supporting modeling techniques and tools (Kettinger, et al., 1997; Aguilar-Saven, 2004). It has become more and more complex to select the right technique and the right tool not only because of the huge range of approaches available but also due to the lack of a guide that explains and describes the concepts involved. In this chapter, we will help readers to bring order to this chaos reviewing several BPM/ISM methodologies and techniques.

Basic Concepts and Foundations

In 1960 Levitt first mentioned the importance of business processes in his important marketing paper (Levitt, 1960). It was not until the last decade, however, that processes have acquired real importance in enterprise design (Aguilar-Saven, 2004). Process-based thinking in the context of organizational change can be regarded as a systems design problem (Earl, 1994; Davenport & Stoddard, 1994). According to the information processing (Tushman & Nadler, 1978) and decision making (Huber & McDaniel, 1986) paradigms of organizational design, processes can be viewed as collections of decision models, each of which is identified by a type of decision and contains a sequence of processing tasks (Moore & Whinston, 1986). These tasks are the smallest identifiable units of analysis, and their optimum arrangement is the critical design variable determining the efficiency of the resulting structures. Therefore, techniques that allow for modeling business process components, experimenting with alternative configurations and process layouts, and comparing diverse proposals for change would be highly suitable for organizational design and business engineering (Giaglis, 2001). Coupled with the widely studied problems of IS design and development, this 'model management' approach presents an opportunity for addressing the 'business/IS fit' problem by means of modeling.

To study and understand systems, one constructs models according to particular viewpoints and using particular modeling technique. Kettinger et al. (1997) present an important overview of methods, techniques, and tools used in Business Process Re-engineering (BPR). As part of that study a list of related business process modeling techniques and tools was published. Davenport (1993) defines process as 'structured, measured sets of activities designed to produce a specified output for a particular customer or market'. The term business process modeling is used to incorporate all activities relating to the transformation

of knowledge about business systems into models that describe the processes performed by organizations (Scholz-Reiter & Stickel, 1996). The term information systems modeling is used in a similar fashion to denote approaches 'seeking to make our abstractions of information systems look more like the real-world systems they represent' (Sol & Crosslin, 1992).

There are many classifications of business processes. Often 'core' and 'supportive' business processes are distinguished. A core (or primary) process is initiated from outside an organization. A supportive (or secondary) process creates the conditions for the primary process to be carried out. This last might be classified in its turn as management processes that control the organization's overall strategies and objectives; and support processes, which support the core processes by offering sufficient resources. Two uses of business process models are distinguished: one for traditional software development, and another to restructure business processes (Phalp, et al., 1999). This suggests a need for different notational approaches, for different modeling purposes and audiences (Phalp, 1998). For the software development process typically diagrammatic notation is typically required, for capturing a legible and understandable view of the business process. Typically the user does not need to play or interact with the model but rather just observe it. When analyzing the business process it is necessary to have more sophisticated mechanisms than qualitative analysis of static diagrammatic models - models that present both dynamic and functional aspects of the process. Finally, in presenting the business process, easily-understood approaches are chosen, and again a readily understandable, typically diagrammatic notation is suitable.

Methodologies are broadly used in software engineering and knowledge engineering (Downs et al., 1998; Wielinga, 1992). In IEEE standard, a methodology is a 'comprehensive, integrated series of techniques or methods creating a general systems theory of how a class of thought-intensive work ought to be performed' (IEEE, 1995). A method is an 'orderly' process or procedure used in the engineering of a product or performing a service (IEEE, 1990). A technique is a 'technical and managerial procedure used to achieve a given objective' (IEEE, 1995).

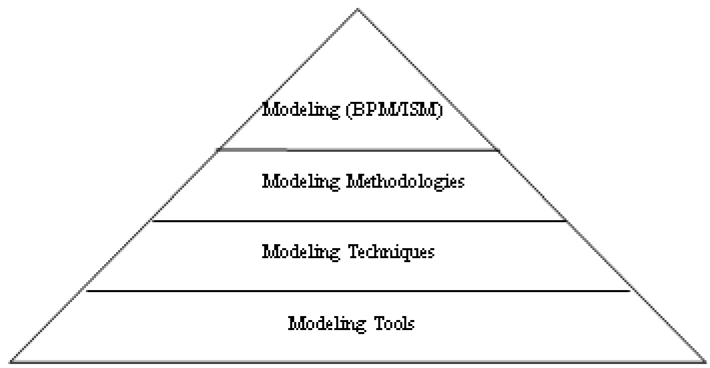
In this chapter, we follow the studies of Giaglis (2001), Aguilar-Saven (2004), Kettinger et al. (1997), to distinguish these concepts as shown in the hierarchic framework for BPM/ISM in Figure 1. According to this decomposition, modeling in general can be thought of as supported by one or more methodologies. Methodology is defined as a collection of problem-solving methods governed by a set of principles and a common philosophy for solving targeted problems. Modeling methodologies are supported by a number of techniques that provide the main analytical focus of our research. Techniques are taken to refer to a set of precisely described procedures for studying and analyzing modeled systems. Specific techniques, as well as their underlying methodologies, can be supported (and, in most cases, are supported) by software modeling tools, such as CASE tools, Workflow Management Systems, process modeling software, and others.

Soft Systems Methodology

Soft Systems Methodology (SSM) is a methodology used to support and to structure thinking about, and intervention in, complex organizational problems. As an approach to organizational process modeling, it can be used both for general problem solving and in the management of change. SSM remains the most widely used and practical application of systems thinking, and other systems approaches such as Critical Systems Thinking have incorporated many of its ideas.

The SSM process consists of seven stages. In these stages, one uses relevant techniques and switches between the real world and the conceptual modeling world where appropriate. These seven stages are:

Figure 1. Modeling methodologies, techniques, and tools for BPM/ISM



1. Appreciating the unstructured problematical situation
2. Understanding the worldviews of the key stakeholders
3. Creating root definitions of relevant systems
4. Making and testing conceptual models based upon worldviews
5. Comparing conceptual models with reality
6. Identifying feasible and desirable changes
7. Acting to improve the problem situation

The seven stages do not represent a single process which can be followed from start to finish, after which a 'right' answer will be obvious. These stages are stages in a process: the process may have to be repeated many times before a reasonable accommodation or agreement can be reached.

The whole process of SSM is a process of mutual learning: the practitioner learns about the organization; the members of the organization learn about the diversity of views about and within their organization, and about their colleagues. The most important site for this learning is in the comparison between conceptually derived models and the real world. When such a comparison is made, the learning gained usually means the model needs to be revised.

At the same time, exposure to the model often changes the problem situation, or at least perceptions of what the problem consists of. Through this conversational process of thinking, discussing, accommodating and re-thinking, practical ways forward may eventually be found.

This methodology helps to understand and analyze a process from the human perspective. One of the techniques used to describe a process is called Rich Pictures, which are highly contextual representations of things. They represent some of the richness of the situation being examined and illustrate issues that will be considered for analysis, reflection and change. They include components such as clients, people involved, tasks performed and environment. This technique is very useful in understanding the interaction of different elements involved in the process and the interaction between processes, although it is not suitable for a structured analysis, or to report a description.

SSADM Methodology

Structured Systems Analysis and Design Method (SSADM) is a systems approach to the analysis and design of information systems. It is not considered as a particular technique for process modeling. It is considered as a set of procedural, technical and documentation standards for systems development. SSADM adopts the waterfall model

of systems development, where each phase has to be completed and signed off before subsequent phases can begin.

The SSADM method involves the application of a sequence of analysis, documentation and design tasks concerned with the following (Downs et al., 1992).

Feasibility Study

Analyze the current situation at a high level. A Data Flow Diagram (DFD) is used to describe how the current system works and to visualize known problems. The following steps are part of this stage:

1. Develop a Business Activity Model
2. Investigate and define requirements
3. Investigate current processing
4. Investigate current data
5. Derive logical view of current services

Requirement Analysis

This stage consists of two parts. The first part is researching the existing environment. In this part, system requirements are identified and the current business environment is modeled. Modeling consists of creating a DFD and LDS (Logical Data Structure) for processes and data structures that are part of the system. In the second part, BSO (Business Systems Options), six business options are presented. One of the options is selected and built.

Requirements Specification

To assist management to make a sound choice, a number of business system options, each describing the scope and functionalities provided by a particular development/implementation approach, are prepared and presented. These options may be supported by technical documentation such as Work Practice Model, LDM (Logical Data Model) and DFD. They also require financial and risk assessments to be prepared, and need to be supported by outline implementation descriptions.

The following steps are part of this stage: Define required system processing; Develop required data model;Derive system functions;Develop user job specifications;Enhance required data model;Develop specification prototypes;Develop processing specification;Confirm system objectives.

Logical System Specification

1. **Logical data design.** In this stage, technically feasible options are chosen. The development/implementation environments are specified based on this choice. The following steps are part of this stage: Define BSOs (Business Systems Options);Select BSO.
2. **Logical process design.** In this stage, logical designs and processes are updated. Additionally, the dialogs are specified as well. The following steps are part of this stage: Define user dialogue;Define update processes;Define enquiry processes.

Physical Design

The objective of this stage is to specify the physical data and process design, using the language and features of the chosen physical environment and incorporating installation standards. The following activities are part of this stage:

1. Prepare for physical design
2. Learn the rules of the Implementation environment
3. Review the precise requirements for logical to physical mapping
4. Plan the approach
5. Complete the specification of functions
6. Incrementally and repeatedly develop the data and process designs

MODELING TECHNIQUES

IDEF Techniques

IDEF was a product of the Integrated Computer-Aided Manufacturing (ICAM) initiative of the United States Air Force (http://www.idef.com), in response to the identification of the need to improve manufacturing operations. 'IDEF' initially stood for 'ICAM DEFinition' language; the IEEE standards recast IDEF as 'Integration DEFinition.'

These 'definition languages' have become standard modeling techniques. They cover a range of uses from function modeling to information, simulation, object-oriented analysis and design and knowledge acquisition, specifically as follows:

1. **IDEF0:** Function Modeling
2. **IDEF1:** Information Modeling
3. **IDEF1X:** Data Modeling
4. **IDEF2:** Simulation Model Design
5. **IDEF3:** Process Description Capture
6. **IDEF4:** Object-Oriented Design
7. **IDEF5:** Ontology Description Capture
8. **IDEF6:** Design Rationale Capture
9. **IDEF7:** Information System Auditing
10. **IDEF8:** User Interface Modeling
11. **IDEF9:** Scenario-Driven IS Design
12. **IDEF10:** Implementation Architecture Modeling
13. **IDEF11:** Information Artifact Modeling
14. **IDEF12:** Organization Modeling
15. **IDEF13:** Three Schema Mapping Design
16. **IDEF14:** Network Design

For business process and information systems modeling, the most useful versions are IDEF0, IDEF1X and IDEF3 which will be explained further below.

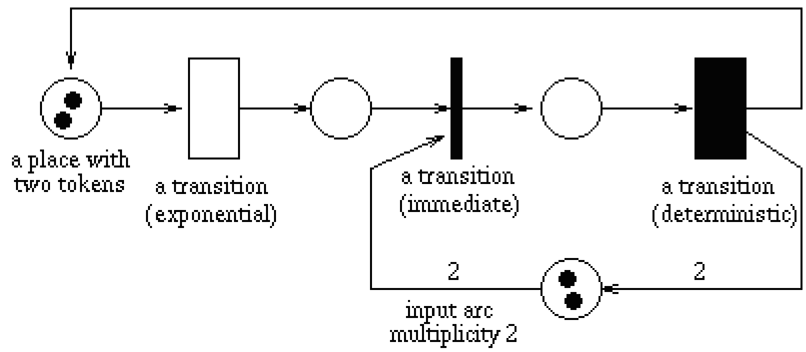
IDEF0 is a modeling technique used for developing structural graphical representations of decisions, actions, and activities of an organization or system. IDEF0 was derived from a well-established graphical language known as the Structured Analysis and Design Technique (SADT). The Air Force commissioned the developers of SADT to develop a function modeling method for analyzing and communicating the functional perspective of a system.

The IDEF0 standard is the most popular process-modeling on the market. The very strict rules in IDEF0 make it suitable for implementation as computer software. By working backwards along the chain from output to input, much data and control can be defined, thus it can be analyzed and improved. The hierarchical structure facilitates quick mapping at a high level. One weakness is the tendency of IDEF0 models to be interpreted as representing a sequence of activities. The activities may be placed in a left to right sequence within decomposition and connected with the flows. It is natural to order the activities left to right because, if one activity's output is used as input by another activity, drawing the activity boxes and concept connections is clearer. Thus, without intent, activity sequencing can be embedded in the IDEF0 model.

IDEF1 is used for information modeling, which captures conceptual views of the enterprise's information. In CIM applications, IDEF1 is generally used to identify what information is currently managed in the organization, identify which of the problems identified during the needs analysis are caused by lack of managing appropriate information, and specify what information will be managed in the 'To-Be' CIM implementation.

IDEF1X is used for data modeling, which captures the logical view of the enterprise's data and is based on an entity relationship model. It was designed as a technique for modeling and analyzing data structures for the establishment of information systems requirements (Mayer et al., 1995). Because it is a design method, IDEF1X is not particularly suited to serve as an AS-IS analysis tool, although it is often used in that capacity as an alternative to IDEF1. IDEF1X is most useful

Figure 2. Example of a Petri Net



for logical database design after the information requirements are known and the decision to implement a relational database has been made. Hence, the IDEF1X system perspective is focused on the actual data elements in a relational database. If the target system is not a relational system, for example, an object-oriented system, IDEF1X is not the best method.

IDEF3 captures precedence and causality relations between situations and events in a form natural to domain experts by providing a structured method for expressing knowledge about how a system, process, or organization works. It allows different views of how things work within an organization. Unlike IDEF0, IDEF3 has been developed for explicitly describing processes. The former shows what is done within the organization while the latter shows how things work with it. From domain experts, descriptions are captured in which the precedence and causality relationships between activities and events of the process are shown. IDEF3 consists of two modeling modes: the process flowed scrimption (PFD), which describes how things actually work in the organization, and the object state transition description (OSTD), which summarizes an object's allowable transitions in a particular process. It is suitable to model both simple and complex processes due to its decomposition ability.

The basic notation of the IDEF3 method consists of a series of square and oblong boxes, and circles and arcs which link them. Attached to each icon is an elaboration form, which contains a description of that icon, reference label, and so on, and a detail of related objects, facts and constrains acting upon it. IDEF3 is used in several areas such as Business Process Engineering (BPE) and Reengineering (BPR), software process definition and improvement, and even in software development and maintenance.

Since they belong to the same family of techniques, IDEF models complement each other effectively and, when combined, can provide a holistic perspective of a modeled system. However, this facility comes at a potentially high complexity of developing and maintaining many different models for a single system.

Petri Nets

Among systems modeling techniques, Petri nets perhaps has received the most attention as a potential candidate for business process modeling (Reising et al, 1992). The concept of Petri nets has its origin in Carl Adam Petri's dissertation submitted in 1962.

A Petri net is one of several graphical and mathematical modeling languages for the description of discrete distributed systems. A Petri net

is a directed bipartite graph, in which the nodes represent transitions (i.e. discrete events that may occur), places (i.e. conditions), and directed arcs (that describe which places are pre- and/or post-conditions for which transitions). As shown in Figure 2, a Petri net consists of places, transitions, and directed arcs that connect them. Arcs run between places and transitions, never between places or between transitions. Input arcs connect places with transitions, while output arcs start at a transition and end at a place. There are other types of arc, for example, inhibitor arcs.

Places may contain any non-negative number of tokens. A distribution of tokens over the places of a net is called a marking. The current state of the modeled system (the marking) is given by the number (and type if the tokens are distinguishable) of tokens in each place. A transition of a Petri net may fire whenever there is a token at the end of all input arcs; when it fires, it consumes these tokens, and places tokens at the end of all output arcs. A firing is atomic, i.e., a single non-interruptible step. The interactive firing of transitions in subsequent markings is called token game

Execution of Petri nets is nondeterministic: when multiple transitions are enabled at the same time, any one of them may fire. The net also doesn't specify whether any of them will fire at all. Since firing is nondeterministic, and multiple tokens may be present anywhere in the net (even in the same place), Petri nets are well suited for modeling the concurrent behavior of distributed systems.

Petri nets are a promising tool for describing and studying systems that are characterized as being concurrent, asynchronous, distributed, parallel, nondeterministic, and/or stochastic. As a graphical tool, Petri nets can be used as a visual-communication aid similar to flow charts, block diagrams, and networks. In addition, tokens are used in these nets to simulate the dynamic and concurrent activities of systems. As a mathematical tool, it is possible to set up state equations, algebraic equations, and other mathematical models governing the behavior of systems.

It has been recognized that basic Petri nets are not succinct and manageable enough to be useful in modeling high-level, complex business processes (Leymann & Altenhuber, 1994). To this end, a number of extensions to the basic Petri net formalism have been proposed. These extensions collectively are referred to as high-level Petri nets and include, for example, generalized stochastic Petri nets (Marsan et al., 1995) and colored Petri nets (Jensen, 1996).

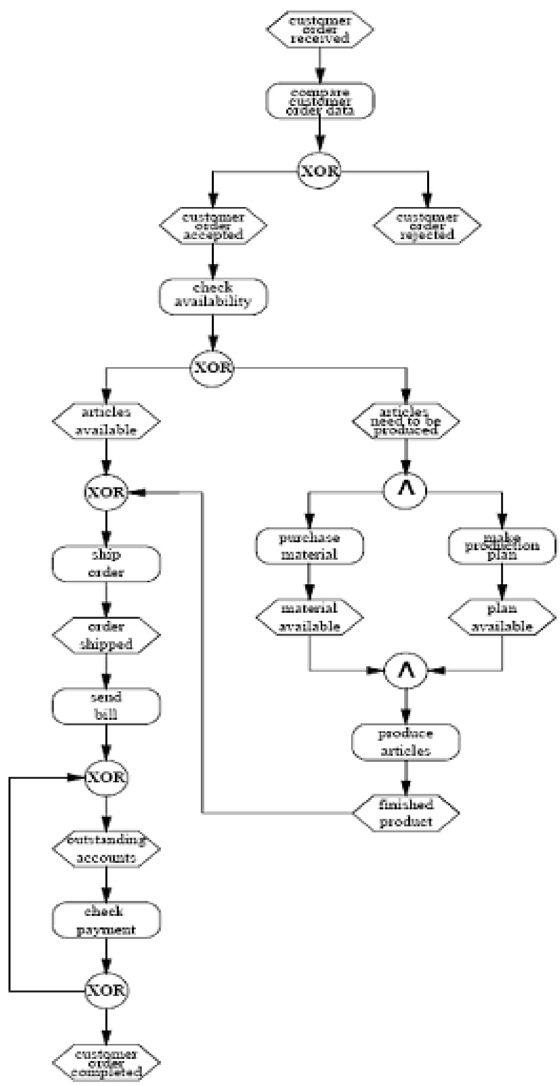
To study performance and dependability issues of systems it is necessary to include a timing concept into the model. There are several possibilities in doing this for a Petri net, but the most common way is to associate a firing delay with each transition. This delay specifies the time that the transition has to be enabled, before it can actually fire. If the delay is a random distribution function, the resulting net class is called stochastic Petri net. Different types of transitions can be distinguished depending on their associated delay; for instance immediate transitions (no delay), exponential transitions (delay is an exponential distribution), and deterministic transitions (delay is fixed).

Event-Driven Process Chain (EPC)

EPC is an intuitive graphical business process description language which is used by many companies for modeling, analyzing, and redesigning business processes. The language is targeted to describe processes on the level of their business logic, not necessarily on the formal specification level, and aims to be easily understood and used by business people.

An EPC provides various connectors that allow the alternative and parallel execution of processes. Furthermore it is specified by the usages of logical operators, such as OR, AND, and XOR. A major strength of EPC is claimed to be its simplicity and easy-to-understand notation. This makes EPC a widely acceptable technique to denote business

Figure 3. Modeling of a business process, using event-driven process chains



processes; for example, the process modeled in Figure 3 models the processing of a customer order, which shows that event-driven process chains are easy to read.

An event-driven process chain consists of the following elements:

Event

Events are passive elements in an EPC. They describe under what circumstances a function or a process works or in which state a function or a process results. In the EPC graph an event is represented as hexagon. In general, an EPC diagram 'must' start with an event and end with an event.

Functions

Functions are active elements in an EPC. A function corresponds to an activity (task, or process step) which needs to be executed. Functions describe transformations from an initial state to a resulting state. Examples of functions are 'capture requirement', or 'check material on stock'. In the EPC graph a function is represented as rounded rectangle.

Logical Connector

In the EPC, the logical relationships between elements in the control flow, that is, events and functions, are described by logical connectors. With the help of logical connectors it is possible to split the control flow from one flow to two or more flows and to synchronize the control flow from two or more flows to one flow. There are three kinds of logical relationships defined in EPC:

Branch/Merge
Branch and merge correspond to making decision about which path to choose among several control flows. A branch may have one incoming control flow and two or more outgoing control flows. When the condition is fulfilled, a branch activates only one of the outgoing control flows and deactivates the others.

The counterpart of a branch is a merge. A merge may have two or more incoming flows and one outgoing control flow. A merge synchronizes activated and deactivated alternatives. The control

will then be passed to the next element after the merge. A branch in the EPC is represented by an opening XOR, whereas a merge is represented as a closing XOR connectors.

Fork/Join

Fork and join correspond to activating all paths in the control flow concurrently. A fork may have one incoming control flow and two or more outgoing control flows. When the condition is fulfilled, a fork activates all of the outgoing control flows in parallel. A join may have two or more incoming control flows and one outgoing control flow. A join synchronizes all activated incoming control flows. In the EPC diagram how the concurrency is achieved is not important. In reality the concurrency can be achieved by true parallelism or by virtual concurrency achieved by interleaving. A fork in the EPC is represented by an opening 'AND', whereas a join is represented as a closing 'AND' connectors.

OR

An 'OR' relationship corresponds to activating one or more paths among control flows. An opening 'OR' connector may have one incoming control flow and two or more outgoing control flows. When the condition is fulfilled, an opening 'OR' connector activates one or more control flows and deactivates the rest of them. The counterpart of this is the closing 'OR' connector. When at least one of the incoming control flows is activated, the closing 'OR' connector will pass the control to the next element after it.

The event-driven process chain shown in Figure 3 is a so-called basic event-driven process chain. It is possible to extend event-driven process chains with entities, business objects, organizational units, and moreover to specify allocation rules and responsibilities.

Organization Unit

Organization units determine which person or organization within the structure of an enterprise is responsible for a specific function. Examples are 'sales department', 'sales manager', 'procurement manager', etc.. An organization unit is represented as an ellipse with a vertical line.

Information, Material, or Resource Object

In the EPC, the information, material, or resource objects portray objects in the real world, for example business objects, entities, and so forth, which can be input data serving as the basis for a function, or output data produced by a function. Examples are 'material', 'order', etc. In the EPC graph such an object is represented as rectangle.

Control Flow

A control flow connects events with functions, process paths, or logical connectors creating chronological sequence and logical interdependencies between them. A control flow is represented as a dashed arrow.

Information Flow

Information flows show the connection between functions and input or output data, upon which the function reads changes or writes.

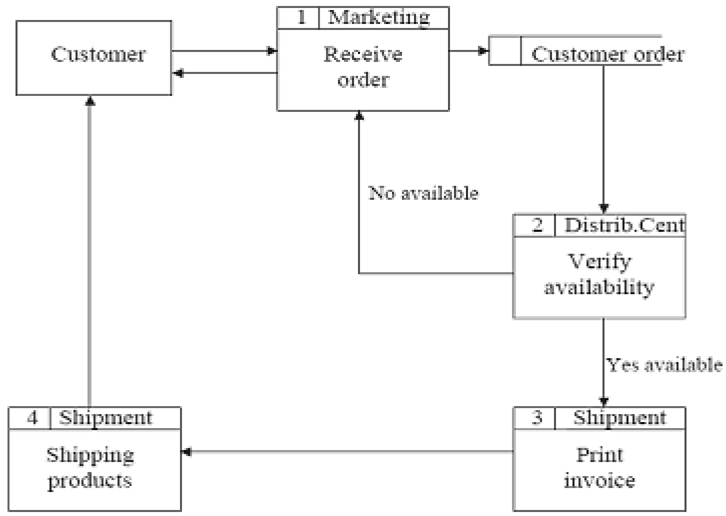
Organization Unit Assignment

Organization unit assignments show the connection between an organization unit and the function for which it is responsible.

Process Path

Process paths serve as navigation aid in the EPC. They show the connection from or to other processes. The process path is represented as a compound symbol composed of a function symbol superimposed upon an event symbol. To employ the process path symbol in an EPC diagram, a symbol is connected to the process path symbol,

Figure 4. Example of a data flow diagram



indicating that the process diagramed incorporates the entirety of a second process which, for diagrammatic simplicity, is represented by a single symbol.

There are a number of tools for creating EPC diagrams, including ARIS Toolset of IDS Scheer AG (http://www.ids-scheer.com/), ADONIS of BOC Group (http://www.boc-group.com/), Visio of Microsoft Corporation (http://office.microsoft.com/visio). Some but not all of these tools support the tool-independent EPC Markup Language (EPML) interchange format. There are also tools that generate EPC diagrams from operational data, such as SAP logs. EPC diagrams use symbols of several kinds to show the control flow structure (sequence of decisions, functions, events, and other elements) of a business process.

Data Flow Diagrams (DFD)

Data flow diagrams (DFD) show how a system is divided into smaller portions and highlight the flow of data between those parts. This context-level DFD describes the processes showing how these processes link together through data stores and how the processes relate to users and the outside world, showing more detail of the system being modeled.

Data flow diagrams were invented by Larry Constantine, the original developer of structured design (Stevens et al., 1974), based on Martin and Estrin's 'data flow graph' model of computation.

Data flow diagrams (DFDs) are one of the three essential perspectives of Structured Systems Analysis and Design Method (SSADM). The sponsor of a project and end users will need to be briefed and consulted throughout all stages of a system's evolution. With a dataflow diagram, users are able to visualize how the system will operate, what the system will accomplish, and how the system will be implemented. The old system's dataflow diagrams can be drawn up and compared with the new system's dataflow diagrams to draw comparisons to implement a more efficient system. Dataflow diagrams can be used to provide the end user with a physical idea of where the data they input ultimately has an effect upon the structure of the whole system from order to dispatch to restock. DFDs are used in discussions between analysts

and users as they can be easily understood and verified, and are easy to draw and amend. Each process can be broken down into sub-processes at a lower level to show more detail. DFD are used in the functional model to specify the meaning of operations and constraints and show functional dependencies. They show how information enters and leave the process, what activities change the information, where information is stored within the process, and the organizational function to which the activity belongs.

Figure 4. shows an example of a data flow diagram. There are 4 key elements in a Data Flow diagram.

Process Entity

The process entity identifies a process taking place. A process must have at least one input (without which it is known as a 'miracle process'), and one output (without which it is known as a 'black hole process'). Both 'miracle' and 'black hole' processes indicate that the DFD is missing a connection between processes and should be re-evaluated (Post & Anderson, 2004). Each process has the following (Dennis, et al., 2005):

1. A Number
2. A Name (verb phrase)
3. A Description
4. At least one input
5. At least one output

Data Flow Entity

The data flow entity identifies the flow of data between processes, data stores and external entities. A data flow cannot connect an external entity to a data source; at least one connection must be with a process. Each data flow has the following (Dennis, et al., 2005):

1. A Name (Noun)
2. A Description
3. One or more connections to a process.

Data Store Entity

The 'Data Store entity' identifies stores of data, both manual and electronic. Each data store has the following (Dennis, et al., 2005):

1. A Number
2. A Name
3. A Description
4. One or more input data flows.
5. One or more output data flows.

External Entity

The external entity identifies external entities which interact with the system; they are usually clients, but can be within the same organization. Each external entity has the following (Dennis et al., 2005):

1. A Name (Noun)
2. A Description

Although they are widely used and have become a standard notation for systems analysis and design, DFDs present a number of limitations (Giaglis, 2001). First, they focus exclusively on data and provide no modeling constructs on which to base representation of work flow, people, events, and other business process elements. Second, they provide no information on decisions and event sequences. Finally, DFDs are static representations of a system and the system's functions that involve data manipulation; therefore, they do not lend themselves easily to analysis or decision making. To facilitate such analysis, data flow diagramming is sometimes complemented by structured textual descriptions of procedures in which data are to be used; these descriptions are called process specifications (Yourdon, 1989).

There are a number of tools for creating DFD diagrams, including ConceptDram, Dia, Micro-

soft Visio, SamartDraw, System Architect, and DFDdeveloper.

Entity-Relationship Diagramming (ER)

An entity-relationship model (ERM) is an abstract conceptual representation of structured data, and is a widely used data modeling technique. Entity-relationship modeling is a relational schema database modeling method, used in software engineering to produce a type of conceptual data model (or semantic data model) of a system, often a relational database, and its requirements in a top-down fashion. Diagrams created using this process are called entity-relationship diagrams, or ER diagrams or ERDs for short. ER diagrams are network models that describe the stored data layout of a system (Yourdon, 1989). Originally proposed in 1976 by Dr. Pin-Shan (Peter) Chen (Chen, 1976), many variants of the process have subsequently been devised.

ER models are usually used during the requirements analysis to describe information needs or the type of information that is to be stored in a database. The data modeling technique can be used to describe any ontology for a certain universe of discourse. In the case of the design of an information system that is based on a database, the conceptual data model is at a later stage (usually called logical design) mapped to a logical data model, such as the relational model; this in turn is mapped to a physical model during physical design.

ER diagrams focus on modeling the data in a system and their interrelationship in a manner entirely independent of the processing that may take place on that data. Such separation of data and operations may be necessitated in cases where the data and their interrelationships are sufficiently complex. For the system analyst, ER diagrams have another advantage: they highlight relationships between data stores in the DFD that otherwise would be visible only in the (textual) process specification (Giaglis 2001).

ER diagrams consist of entities and the relationships between entities. An entity may be defined as a thing which is recognized as being capable of an independent existence and which can be uniquely identified. An entity is an abstraction from the complexities of some domain. When we speak of an entity we normally speak of some aspect of the real world which can be distinguished from other aspects of the real world (Beynon-Davies, 2004). Entities can be thought of as nouns, for example a computer, an employee, a song, a mathematical theorem. Entities are represented as rectangles.

A relationship captures how two or more entities are related to one another. Relationships can be thought of as verbs, linking two or more nouns. Relationships are represented as diamonds, connected by lines to each of the entities in the relationship.

For business process modeling, ER diagrams focus too much on data and their interrelationships and hence provide no constructs for modeling other process elements. Even more important, they provide no information about the functions depicted that create or use these data. Finally they are entirely static representations, providing no time-related information that could drive analysis and measurement (Giaglis, 2001).

There are a number of tools for creating ER diagrams, including Ferret, BerModelo, ModelRight, MySQL Workbench, Open System Architect, and StarUML.

Unified Modeling Language (UML)

Introduced in 1997 and supported by major industry-leading companies, the unified modeling language (UML) was rapidly accepted throughout the object-technology community as the standard graphical language for specifying, constructing, visualizing, and documenting software-intensive systems (Booch et al, 1999). UML (http://www.

uml.org) is officially defined at the Object Management Group (OMG) (http://www.omg.org) by the UML metamodel, a Meta-Object Facility metamodel (MOF).

The notation used by the UML is graphical in nature, easy to master and, for the most part, simple to understand. Although some people claim that the UML has too many diagrams (13 types of diagram), in reality there are only three basic types. Six diagram types represent static application structure, three represent general types of behavior, and four represent different aspects of interactions:

1. Structure Diagrams include the Class Diagram, Object Diagram, Component Diagram, Composite Structure Diagram, Package Diagram, and Deployment Diagram.
2. Behavior Diagrams include the Use Case Diagram (used by some methodologies during requirements gathering); Activity Diagram, and State Machine Diagram.
3. Interaction Diagrams, all derived from the more general Behavior Diagram, include the Sequence Diagram, Communication Diagram, Timing Diagram, and Interaction Overview Diagram.

UML attempts to address this gap by being a 'universal' language, covering everything from business process representation to database schema depiction and software components modeling. According to its developers, UML will reduce the degree of confusion within the industry surrounding modeling languages. Its adoption would settle unproductive arguments about method notations and model interchange mechanisms, and would allow the industry to focus on higher leverage, and more productive activities (UML, 1997).

UML is targeted mostly to systems modeling, although an 'extension for business modeling' has also been developed (Giaglis, 2001). There are a number of tools for creating UML diagrams, including non-commercial or free UML tools, such as Acceleo (http://www.acceleo.org/), ArgoUML (http://argouml.tigris.org/), BoUML(http://bouml.free.fr/), Eclipse(http://www.eclipse.org/), NetBeans(http://www.netbeans.org/), and commercial or proprietary UML tools, such as Rational Rose(http://www.ibm.com/), SmarDraw(http://www.samrtdraw.com/), Cadifra UML Editor(http://www.cadifra.com/), Microsoft Visio(http://office.microsoft.com), and others. Readers can refer to UML Forum (http://www.uml-forum.com) and Wikipedia (http://en.wikipedia.org/wiki/List_of_UML_tools) to get more information on UML tools.

CASE STUDY: CONCEPTUAL MODELING OF COLLABORATIVE MANUFACTURING FOR CUSTOMIZED PRODUCTS

In this section, we will present a case study about business process and information systems modeling. It is based on our previous work (Yan, et al., 2008).

Background and Motivation

Today's market environment is characterized by diverse customer tastes and preferences. People are no longer willing to sacrifice their preferences but are looking for exactly what they want and need (Pine, 1993). Advanced manufacturing systems are evolving towards a more agile environment that can make quick responses to the changing market environment and customer requirements.

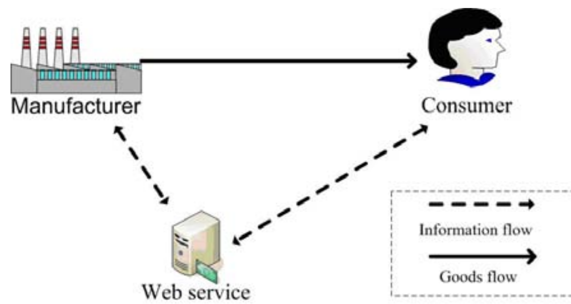
Mass customization, relating to the ability to provide tremendous variety and individual customization (Pine, 1993), is implemented by many companies to gain a significant competitive advantage (Kotha, 1995). Some researchers argue that mass customization should provide customers with whatever they want, whenever they want it, wherever they want it and however

they want it. As customers and their needs grow increasingly diverse, however, the degree of product variety is controlled prudently in industry practice (Moozakis, 2002). This phenomenon is caused by the limitations of the traditional value chain and production paradigm of mass customization (Zipkin, 2001). Researchers have suggested that the problems lie in both the supply side (i.e. the concerns of scale economies and costs mentioned in Randall & Ulrich 2001) and the demand side (i.e. the problem of variants cannibalization stated in Hui 2004; Moorthy, 1984). In addition, transformations and revolutions are carried out to improve mass customization by adopting advanced information technologies (i.e. Turowski, 2002).

With the development of manufacturing systems, some advanced manufacturing systems and concepts are introduced, such as the manufacturing grid, virtual enterprise, and holonic manufacturing (Camarinha-Matos & Afsannanesh, 1999; Ding, et al, 2008). These technologies are improving the manufacturing efficiency to suit the changing market environment and customer requirements. Taking advantage of these advanced technologies to facilitate mass customization and improve the customer-perceived value of mass customization raises a challenging issue. Having many common and even complementary characteristics, these concepts are not necessarily contradictory. Combinations of these approaches are possible, and may help to achieve collaborative intelligent manufacturing for a more successful mass customization.

In this section, we intend to explore how such collaborative manufacturing can better solve the problem of product variety to improve mass customization. First, we will describe a collaborative manufacturing process for customized production, after which the corresponding conceptual model of collaborative intelligent manufacturing for customized product is presented. A description and conceptual modelling of such collaborative process could provide deeper understanding of the application of collaborative manufacturing in personalized production and facilitate early detection and correction of system development errors.

Figure 5. The traditional value train of mass customization



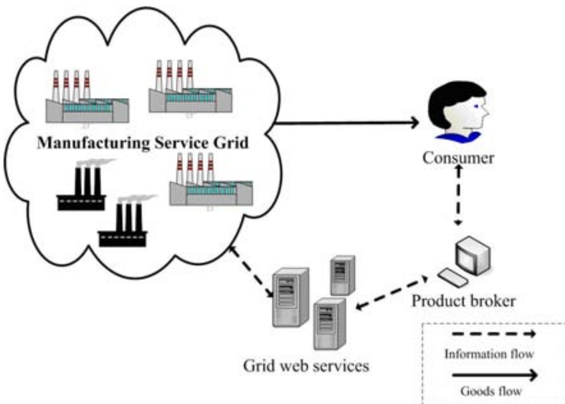
Modeling

Collaborative Manufacturing Process

A scenario will be used to demonstrate how collaborative manufacturing facilitate mass customization and improve the customer-perceived value of mass customization.

Let us suppose a customer ('CT') is interested in acquiring a customized bicycle. Figure 5 shows a traditional value chain of mass customization. The producer provides a set of product or component variants on the web service. When the consumer finishes his product configuration on the web configuration system, the manufacturer produces the product and delivers it to the consumer. Within the traditional value chain of mass customization, consumer could do the product configuration according to their preferences. However, the producer cannot provide a full range of product variety because of costs and cannibalization effects. The consumers may only choose a limited set of frame types and colors for his customized bicycle. For example, CT would like his customized bicycle be made with a 'Titanium TY12' frame, and to be painted with 'ColorType32'. If

Figure 6. Collaborative manufacturing for customized production



there is no manufacturer providing both 'Titanium TY12' and 'ColorType32', CT has to sacrifice his preference and choose another options offered by the manufacturer.

In collaborative manufacturing, each producer focus on their core competence, in order to minimize their production cost and market mediating costs. A typical collaboration manufacturing process for customized production is shown in Figure 6.

The major components are the manufacturing service grid, grid web services, and the product broker. The manufacturing service grid is composed of manufacturers all over the world, which focuses on their individual core manufacturing competences and offers special manufacturing services. Information about these manufacturing services is all encapsulated in the grid web services, with a well-defined interface using WSDL, a standardized messaging protocol such as SOAP, and a service address that a requester can use to access the service. The grid web services can be accessed and invoked programmatically by software agents. The product broker is such a software agent that can help consumers to customize their product and find a corresponding manufacturing service.

With this novel value chain, we have the following scenario and transaction:

Stage 1: Product Development Verification
The customer 'CT' wants his customized bicycle, so he asks for help from the product broker. With its domain knowledge, the product broker tells CT the three generic production tasks required in bicycle production, namely frame fabrication, frame painting, and bicycle assembly; these tasks must be performed in this order. Frame fabrication consists of cutting and welding tubes into unfinished frames. Frame painting consists of adding color to frames and applying decals and a final clear finish. In assembly, components such as wheels, tires, suspension, and drive trains are attached to the finished frame. Firms need not collocate operations.

The consumers might add some additional production tasks to meet his individual needs. In this scenario, we assume that CT accepts the three production tasks and does not request any more additional production tasks.

Stage 2: Product Broker Self Development
After the product development verification, the product broker delegates to three sub-brokers: the frame broker, painting broker and assembly broker. All three sub-brokers are responsible to the bicycle production broker, and will search the corresponding manufacturing services.

Stage 3: Manufacturing Services Verification
In searching and finding the corresponding manufacturing services within the manufacturing grid services, the product broker will offer a set of available manufacturing services for frame fabrication, painting and assembly. The customer CT could either choose the services himself according to his preferences or delegate the product broker to choose the preferred services by the agent's negotiation. In this scenario, we assume CT chooses manufacturer M1 which is the only frame provider of the 'Titanium TY12' frame to do the frame fabrication, manufacturer M2 who provides painting of ColorType32 to paint the

frame, and Manufacturer M3 to do the assembly work.

Stage 4: Production and Delivery
After the completion of the manufacturing services verification, M1, M2, M3 will carry out their manufacturing services in order. The physical distribution network (such as UPS) will be responsible for the transportation between M1, M2 and M3. After the production, the finished customized bicycle will ship to CT.

The main difference between traditional mass customization and collaboration manufacturing for customized production is that the manufacturing service which the latter offers is separated into several subtasks and is finished by the collaboration of corresponding manufacturing services, while the production of the former is dominated by an individual manufacturer. The features of collaborative manufacturing could offer the following benefits:

1. Producers could focus on their core competencies. By efficiently producing its core product, the producer could reduce both production cost and the response time.
2. Without having to handle other aspects of business activities, the producers could reduce market mediation costs.
3. The customer control the order-to-delivery time to suit his need by agent's negotiation with corresponding manufacturing services.
4. Consumers, as innovators, would achieve greater freedom to 'design' their product. Not only could more product variety be produced by collaborative manufacturing, but customers could even design their products by adding new production tasks.
5. The interaction of the autonomous agent and services could lower co-ordination costs.

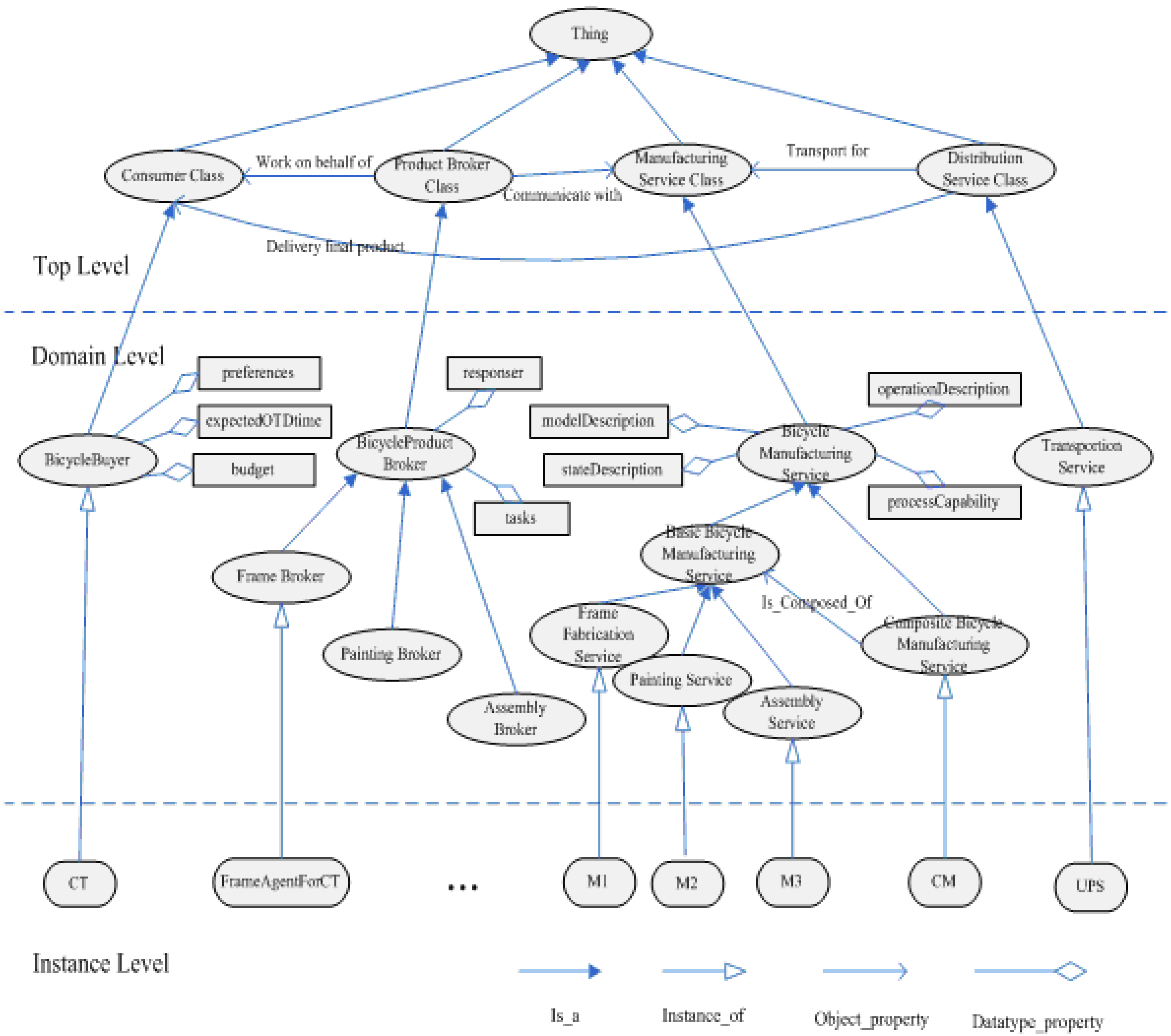
Conceptual Modeling

Within the information systems field, the task of conceptual modeling involves building a representation of selected phenomena in the domain (Wand & Weber, 2002). The model built in this section necessarily entails or embodies some sort of world view with respect to a given domain. The world view, referred as a conceptualisation, is often conceived as a set of concepts (e.g. entities, attributes, and processes), their definitions and their inter-relationships (Uschold & Gruninger, 1996).

Concepts Organization
The conceptual model of collaborative manufacturing for customized production has been designed to model the foundation for collaborative manufacturing applications, which have been captured in four key based classes: Consumer Class, Product Broker Class, Manufacturing Service Class, and Distribution Service Class. A portion of semantic schema of the collaborative manufacturing for customized production is shown in Figure 7. The subset of the overall collaborative manufacturing schema is sufficient to demonstrate the model. Because of the complexity of this figure in mind, many links, such as Is_a, Instance_of, Object_property and Dataype_property, have been omitted.

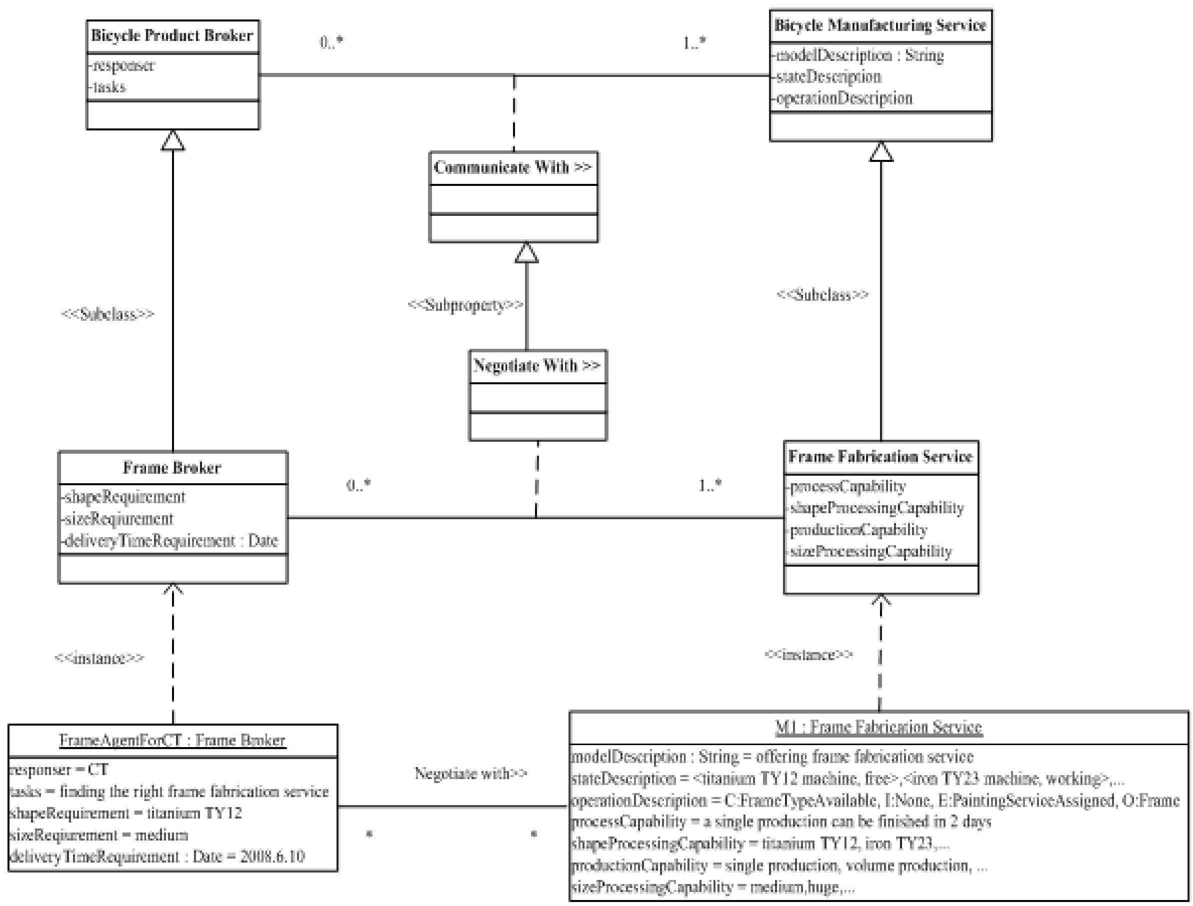
The model which represents collaborative manufacturing for customized production is produced in three levels: Top Level, Domain Level, and Instance Level. The entities at the instance level correspond to the instances of domain classes, while the domain classes inherit the attributes from top level classes. For example, the object property of finalGoodDelivery (omitted in Fig. 6-7) at the instance level is an instance of a domain level object property 'delivery final product', which in turn inherits from the top level object property.

Figure 7. A partial schema



The datatype property of a manufacturing service class consists of an optional modelDescription, a stateDescription and an operationDescription. The modelDescription describes the general information about the service type, while the stateDescription defines the internal state maintained by the manufacturing service. The operationDescription is a 4-tuple <I, O, C, E> (Kumar et al, 2006). I and O represent the data elements accepted by the service during invocation and made available after the invocation of this operation respectively. C is the set of conditions that should be true for this operation to be invoked. E is a set of expressions that become true after the invocation of this operation. For the painting service in the scenario, the <I, O, C, E> in simplified form would be <Frame, ColoredFrame, ColorTypeAvailable, AssemblyServiceAssigned>. The datatype property and object property of class have the features of inheritance and polymorphism, which we will further discuss in the next section.

Figure 8. Inheritance and polymorphism of the manufacturing services



Manufacturing Service Customization

The characteristics of inheritance and polymorphism in the model can be used to provide differentiated and customized manufacturing service to different client. As manufacturers focus on their core competences, the manufacturing services they provided would evolve to incorporate changing requirements. Interface inheritance can be effectively applied to enable different clients of same service to experience different behavior. For instance, the manufacturer M1 in the scenario is a bicycle manufacturer whose core competence is its frame fabrication technology. It can provide the shape of 'Titanium TY12' frame, which other bicycle manufacturers cannot provide. Corresponding to the domain level, the frame fabrication service is modelled with the inheritance and polymorphism from the bicycle manufacturing service, which is shown in Figure 8. The frame agent for CT, which is an instance of Frame Broker, would negotiate with the M1 service for the frame fabrication service.

Given a base Manufacturing Service MS_{base} and the derived Manufacturing Service $MS_{derived}$', $MS_{derived}$ may add new properties to the set of properties inherited from MS_{base}. For instance, in Figure 8, there are new properties of processCability, shapeProcessingCapability, productionCapability, sizeProcessingCapability in the Frame Fabrication Service. On the other hand, the derived service can maintain additional state elements (such as

the specific machine working state), apart from the state inherited from the base service.

Conclusion

In this session, a case has been used to demonstrate the business process and information systems modeling. A conceptual model has been described via a customized bicycle buying scenario. This model exhibits several features of the collaborative manufacturing process, which would benefit both producers and customers.

1. Separation of production tasks enables producers to focus on their core competencies and efficient production, reducing the production and operating cost, market mediation costs and response time.
2. The degree of product variety is enlarged by collaborative manufacturing, and the customer has greater freedom to customize their product.
3. Mediated by autonomous agent and services, the co-ordination costs of the transaction are reduced. Furthermore, customer could achieve a more satisfying order-to-delivery time through the agent's negotiation.

The conceptual model of the collaborative manufacturing process is important, because it only provides a better understanding of how collaborative manufacturing could improve mass customization, but also shows how early detection and correction of system development errors could be facilitated. This could be further developed to serve as a foundation for an architecture enabling information integration of collaborative manufacturing. The development of such a formal conceptual model provides the basis for formal study of collaborative manufacturing for customized products.

SUMMARY

In this chapter, we first gave a brief introduction to the background and the importance of business process modeling and information systems modeling. The basic concepts of information systems and business process modeling were introduced. After presenting a hierarchic framework for IS/BPM, several IS/BPM methodologies and techniques were reviewed to give a better understanding of information systems and business process modeling. Their corresponding development tools were also introduced. A case has been presented to demonstrate a demonstration of the business process and information systems modeling.

Business process and information systems modeling are a much-researched field, and the span of the research agenda is still being formulated. There are many researches relates to IS/BPM which will benefit readers for further readings. Within the field of knowledge representation research, for example, there are many emerging researches on ontology research and knowledge level modeling (Uschold, 1998; Corcho, et al, 2003). With the development of the semantic web, OWL also becomes popular as a ontology language for a formal model in the semantic web. In the field of requirement engineering, Tropos, which is founded on concepts used to model early requirements, is proposed for requirements-driven information systems engineering (Castro et al, 2002; Giunchiglia et al, 2002).

REFERENCES

Aguilar-Saven, R. S. (2004). Business process modeling: Review and framework. *International Journal of Production Economics*, *90*(2), 129–149. doi:10.1016/S0925-5273(03)00102-6

Beynon-Davies, P. (2004). *Database Systems*. Basingstoke, UK: Palgrave.

Booch, G., Rumbaugh, J., & Jacobson, I. (1999). *Unified Modeling Language User Guide*. Reading, MA: Addison-Wesley.

Camarinha-Matos, L. M., & Afsarmanesh, H. (1999). *Infrastructure for Virtual Enterprise*. Norwell, MA: Kluwer Academic Publisher.

Castro, J., Kolp, M., & Mylopoulos, J. (2002). Towards requirements-driven information systems engineering: the Tropos project. *Information Systems*, *27*, 365–389. doi:10.1016/S0306-4379(02)00012-1

Chen, P. P. (1976). The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems*, *1*(1), 9–36. doi:10.1145/320434.320440

Corcho, O., Fernandez-Lopez, M., & Gomez-Perez, A. (2003). Methodologies, tools and languages for building ontologies: Where is their meeting point? *Data & Knowledge Engineering*, *46*(1), 41–64. doi:10.1016/S0169-023X(02)00195-7

Davenport, T. H. (1993). *Process Innovation: Reengineering Work through Information Technology*. Boston: Harvard Business School Press.

Davenport, T. H., & Sort, J. E. (1990). The new industrial engineering: Information technology and business process redesign. *Sloan Management Review*, *31*(4), 11–27.

Davenport, T. H., & Stoddard, D. B. (1994). Reengineering: Business change of mythic proportions? *Management Information Systems Quarterly*, *18*(2), 121–127. doi:10.2307/249760

Dennis, A., Wixom, B., & Roth, R. (2005). *Systems Analysis and Design* (3rd ed.). Mahwah, NJ: Wiley.

Ding, Y., Tao, F., Sheng, B., & Zhou, Z. (2008). Modelling and application of optimal-selection evaluation for manufacturing grid resource. *International Journal of Computer Integrated Manufacturing*, *21*(1), 62–72.

Downs, E., Clare, P., & Coe, I. (1992). *Structured Systems Analysis and Design Method: Application and Context*. Hertfordshire, UK: Prentice-Hall.

Earl, M. J. (1994). The new and the old of business process redesign. *The Journal of Strategic Information Systems*, *3*(1), 5–22. doi:10.1016/0963-8687(94)90003-5

Giaglis, G. M. (2001). A taxonomy of business process modeling and information systems modeling techniques. *International Journal of Flexible Manufacturing Systems*, *13*(2), 209–228. doi:10.1023/A:1011139719773

Giunchiglia, F., Mylopoulos, J., & Perini, A. (2002). The Tropos software development methodology: Processes, models and diagrams. In *Proceedings of the 1st. Int. Conference on Autonomous Agents & Multi-Agent Systems (AAMAS 2002)*, (pp. 35–36), Bologna, Italy. New York: ACM Press.

Huber, G. P., & McDaniel, R. R. (1986). The decision making paradigm of organisation design. *Management Science*, *32*(5), 576–589. doi:10.1287/mnsc.32.5.572

Hui, K. (2004). Product variety under brand influence: An empirical investigation of personal computer demand. *Management Science*, *50*(5), 686–700. doi:10.1287/mnsc.1030.0200

IEEE. (1990). *IEEE Standard Glossary of Software Engineering Terminology, Std. 610.12–1990*. Washington, DC: IEEE Computer Society.

IEEE. (1995). *IEEE Guide for Software Quality Assurance Planning, Std. 730.1–1995*. Washington, DC: IEEE Computer Society.

IEEE. (1996). *IEEE Standard for Developing Software Life Cycle Processes. Std. 1074–1995*. Washington, DC: IEEE Computer Society.

Jensen, K. (1996). *Colored Petri Nets: Basic Concepts, Analysis Methods and Practical Use*. Berlin: Springer-Verlag.

Kettinger, W. J., Teng, J. T. C., & Guha, S. (1997). Business process change: A study of methodologies, techniques, and tools. *Management Information Systems Quarterly*, *21*(1), 55–80. doi:10.2307/249742

Kotha, S. (1995). Mass customization: Implementing the emerging paradigm for competitive advantage. *Strategic Management Journal*, *16*, 21–42. doi:10.1002/smj.4250160916

Kumar, A., Neogi, A., & Ram, J. D. (2006). An OO based semantic model for service oriented computing. *IEEE International Conference on Services Computing (SCC'06)*, (pp. 85-93).

Levitt, J. (1960). Marketing myopia. *Harvard Business Review*, *38*(4), 45–56.

Leymann, F., & Altenhuber, W. (1994). Managing business processes as an information resource. *IBM Systems Journal*, *33*(2), 326–348.

Marsan, (1995). *Modeling with Generalised Stochastic Petri Nets*. Chichester, UK: Wiley.

Mayer, R. J., Benjamin, P. C., Caraway, B. E., & Painter, M. K. (1995). A framework and a suite of methods for business process reengineering . In Grover, V., & Kettinger, W. J. (Eds.), *Business Process Change: Concepts, Methods and Technologies* (pp. 245–290). Harrisburg, PA: Idea Group Publishing.

Moore, T. C., & Whinston, A. B. (1986). A model of decision making with sequential information acquisition. *Decision Support Systems*, *2*(4), 289–308. doi:10.1016/0167-9236(86)90001-1

Moorthy, K. S. (1984). Market segmentation, self-selection, and product line design. *Marketing Science*, *3*(4), 288–307. doi:10.1287/mksc.3.4.288

Moozakis, C. (2002). Nissan wants to be like Dell. *Internet Week, 11*.

Orman, L. V. (1995). A model management approach to business process reengineering. In *Proceedings of the American Conference on Information Systems*, Pittsburgh, August 1995.

Phalp, K. T. (1998). CAP framework for business process modeling. *Information and Software Technology*, *40*(13), 731–744. doi:10.1016/S0950-5849(98)00058-5

Phalp, K. T. (1999). Quantitative analysis of static models of processes. *Journal of Systems and Software*, *52*(2), 105–112. doi:10.1016/S0164-1212(99)00136-3

Pine, B. J. (1993). *Mass Customization: The New Frontier in Business Competition*. Boston: Harvard Business School Press.

Post, G. V., & Anderson, D. L. (2004). *Management Information Systems* (3rd ed.). Boston: Irwin, McGraw-Hill.

Randall, T., & Ulrich, K. (2001). Product variety, supply chain structure, and firm performance: Analysis of the U.S. bicycle industry. *Management Science*, *47*(12), 1588–1604. doi:10.1287/mnsc.47.12.1588.10237

Reising, W., Muchnick, S. S., & Schnupp, P. (Eds.). (1992). *A Primer in Petri Net Design*. Berlin: Springer-Verlag.

Scholz-Reiter, B., & Stickel, E. (Eds.). (1996). *Business Process Modeling*. Berlin: Springer-Verlag.

Sol, H. G., & Crosslin, R. L. (Eds.). (1992). *Dynamic Modeling of Information Systems, II*. Amsterdam: North-Holland.

Stevens, W., Myers, G., & Constantine, L. (1974). Structured design. *IBM Systems Journal*, *13*(2), 115–139.

Turowski, K. (2002). Agent-based e-commerce in case of mass customization. *International Journal of Production Economics*, *75*(1-2), 69–81. doi:10.1016/S0925-5273(01)00182-7

Tushman, M. L., & Nadler, D. A. (1978). Information processing as an integrating concept in organization design. *Academy of Management Review*, *3*(3), 613–624. doi:10.2307/257550

Uschold, M. (1998). Knowledge level modelling: concepts and terminology. *The Knowledge Engineering Review*, *13*(2), 5–29. doi:10.1017/S0269888998001040

Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, *11*(2), 93–136. doi:10.1017/S0269888900007797

Wand, Y., & Weber, R. (2002). Research commentary: Information systems and conceptual modeling – a research agenda. *Information Systems Research*, *13*(4), 363–376. doi:10.1287/isre.13.4.363.69

Wielinga, B. J., Schreiber, A. T., & Breuker, J. A. (1992). KADS: a modeling approach to knowledge engineering. *Knowledge Acquisition*, *4*(1), 5–53. doi:10.1016/1042-8143(92)90013-Q

Yan, J. Q., Ye, K., Wang, H. Q., & Hua, Z. S. 2008. Conceptual Modeling of Collaborative Intelligent Manufacturing for Customized Products: An Ontogical Apporach. In W. Huang and H.H. Teo (eds.), *12th Pacific Asia Conference on Information Systems (PACIS 2008),* (pp. 1446-1456). Suzhou, China: City University Hong Kong.

Yourdon, E. (1989). *Modern Structured Analysis*. Englewood Cliffs, NJ: Prentice-Hall International.

Zipkin, P. (2001). The limits of mass customization. *MIT Sloan Management Review*, *42*(3), 81–87.

Chapter 7
Sensor Integration and Data Fusion Theory

In previous chapters, the engineering scientific foundations of manufacturing intelligence (such as the knowledge-based system, Multi-Agent system, data mining and knowledge discovery, and computing intelligence) have been discussed in detail. Sensor integration and data fusion is another important theory of manufacturing intelligence.

With the development of integrated systems, there is an urgent requirement for improving system automaticity and intelligence. Without improvement, the complexity and scale of systems are increased. Such systems need to be more sensitive to their work environment and independent state, and obviously, single sensor technology hardly meets these requirements. Multi-sensor and data fusion technology are therefore employed in automatic and intelligent manufacturing as it is more comprehensive and accurate than traditional single sensor technology if the information redundancy and complementarity are used reasonably. In theory, the outputs of multi-sensors are mutually validated. Multi-sensor integration is a brand new concept for intelligent manufacturing, and without doubt, sensor integration-based intelligent manufacturing is the development orientation of manufacturing in the future.

DOI: 10.4018/978-1-60566-864-2.ch007

With reference to the information fusion problem of the multi-sensor integration system, the development state, technical background, application scope and basic meaning of the multi-sensor integration and the data fusion are first reviewed in this chapter. Secondly the classification, level, system structure and function model of the data fusion system is discussed. The theoretical method of the data fusion is then introduced, and finally, attention is paid to cutting tool condition detection, machine thermal error compensation and online detection and error compensation because those are the main applications of multi-sensor data fusion technology in intelligent manufacturing.

INTRODUCTION

Background and Development

Multi-sensor integration is a system-controlled process based on integrating information from

different sensors when a multi-sensor system completes a task. It emphasizes the conversion of different data and the overall flow structure for the system. Multi-sensor fusion is a specific stage in the process of multi-sensor integration, in which sensor information is merged into unified comprehensive information. The specific methods and procedures of data conversion and merger are emphasized.

Multi-sensor integration and fusion technology is actually a comprehensive technology of multi-source information. In multi-sensor integration and fusion, the most consistent estimate of a measured object and its nature are obtained through analysis and comprehension for data information from various sensors.

Multi-sensor fusion research started as early as 1973, when research institutions in the United States, funded by the Department of Defense, started to study sonar signal understanding systems. From then on, multi-sensor fusion technology developed rapidly. Apart from being used in the C^3I (Command/Control/Communication Intelligence) system, many kinds of sensors are now widely used to gather information in industrial control, robotics, air traffic control, marine surveillance and management Multi-sensor fusion research has become an issue of concern in the military, production and high-technology development areas.

In 1988, the U.S. Department of Defense listed the C^3I multi-sensor fusion technology as one of 20 key technologies of research and development in the 1990s. The C^3I expert group led by the U.S. Secretary of Defense set up a specialized data fusion group and organized a number of thematic plans to study multi-sensor fusion technology, and since 1992 some 100 million U.S. dollars annually have been invested in multi-sensor fusion technology. Related journals are also published and specialized conferences are held internationally, so that there is a well researched environment for multi-sensor fusion technology. In October 1994, the first international conference on multi-sensor fusion and integration for intelligent systems, launched by IEEE, was held in Las Vegas (Dasarathy B.V., 1997). It signals that research and application of multi-sensor integration and information fusion have come into prominence.

Multi-sensor fusion theory is divided into numerical approach research of similar information fusion, and symbol approach research of information integration of different types. Numerical approach research of similar information fusion, especially distribution of all kinds of optimal, sub-optimal or part dispersed algorithms, is more dominant, while symbol approach research of information integration of different types is more difficult in theoretical research mainly oriented to exploratory research. There are many application systems with the feature of multi-sensor information integration, such as the early warning integrated navigation system, using INS, or Inertial Navigation System and the Global Positioning System (GPS) to improve accuracy by multi-sensor data fusion.

Basic Principles and Its Process

Basic Integrated Principles

Because design requirements for individual systems vary, different forms are used for the process of multi-sensor integration, although some basic principles are shared by many of the implementation processes. Figure 1 presents the framework of multi-sensor integration which is composed of these basic principles. A certain-process is detected by many sensors. Suppose the number of sensors is n. Here n is a variable. Before integration, each sensor data is modeled effectively. Sensor models represent the uncertainty and error of each sensor data, usually assuming that the uncertainty of sensor data is subject to the Gaussian distribution. After sensor data modeling, three different perception treatments can be applied, such as fusion, separate operation and guiding or cueing to integrating.

Figure 1. Functional diagram of multi-sensor integration and integration

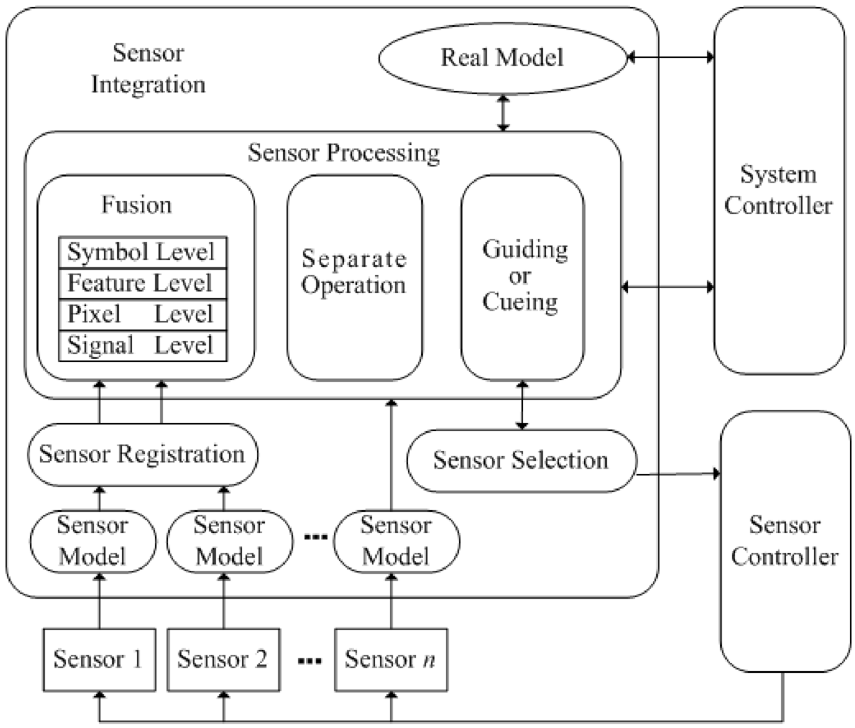


Figure 1 shows the data fusion process for the information obtained from sensor 1 and sensor 2. Before integration, each sensor data must be registered. Sensor registration takes the form of measures (such as geometric transformation) which are employed to unify the dimensions in space and time coordinates of each sensor data; thus, the integrated data have the same time and space dimensions. If the data among sensors varies greatly, the work of other sensors will be indirectly affected. Other sensors will then be impacted indirectly by separate operation of this sensor through impacting on the system controller and true model. Guiding or cueing of perception treatments means that the work of other sensors is guided or cued by a sensor data. This typical example of multi-sensor integration is adopted in many robot applications, such as the work of a tactile array installed in the end of actuator being guided through using visual information.

The result of the perception treatments is considered as the input of the actual model. The actual model is used to store environmental information about system work. The stored environmental information includes prior information and perception information, freshly obtained. In the process of the highest level reasoning, the actual model is used to reason, and the follow-up treatment of perceptual information and the work of the system controller are guided through the results of processing. According to different application needs, the information stored in the actual model takes different forms. For example, in target identification, the actual model includes objectives expression which is only identified by system, while in mobile robot navigation; the actual model includes all the expressions of the robot's work

environment (such as environmental objective characteristics and topographical features). Sensor choice consists of certain measures taken to select or configure sensor groups used by the system. The selection process is accomplished both in the initial system design stage and in the practical work stage of sensors. In the work process, once the environment changes or the state of systems changes (as a result of sensor failure, for instance), the work process is used to determine the most appropriate sensor or sensor group to guide the other sensors to work.

Integrated Approach

Logical Sensor

The concept of logical sensors was put forward by Henderson and Silcrat (1984). It is an abstract definition of sensors, employed to provide an integrated unified structure of sensors (Henderson & Hansen, 1985). Multi-sensor systems have both portability and the ability to adapt to technological change by using logical sensors (Figliolini et al., 2000).

In a logical sensor networks, each logical sensor is looked upon as an element. The logic of sensors has been defined by the 'Name of Logical Sensors'. 'Output Feature Vector' describes data types of output vector flow produced by logical sensors; 'Controlling Orders' are inputted to the logical sensors, including orders needed by the control logical sensors, and including orders used to transmit to other lower level sensors of the network; 'Explanation of Control Orders' accepts control orders, and appropriate orders are then sent to the logical sensors of the lower level in the network. The role of 'Selector' is to monitor control orders which are sent to the logical sensors and outcomes of all kinds of the 'Procedure Unit' (that is to say, 'Selector' plays the role of a 'small expert system'). Each procedure unit has the responsibility for completing all the numerical calculation which is entered into the unit and the input of the logical sensors is the output vector of the lower level logical sensors in the network.

Artificial Neural Network

A robust architecture is provided by the artificial neural network; by using it, a model is established in the process of multi-sensor integration.

- **Multi-perception:** Differing from the majority of statistical pattern recognition methods, it is not necessary to assume the distribution of sensor data when using multi-sensor, as such as data are of non-Gaussian distribution. As long as there is a sufficient number of training samples and a sufficiently long training time, the sensor data are classified through learning of multi-layer perception.
- **Association Storage:** Neuron after training is used to express sensor information, through association recall, the complex mix of neurons can respond to different sensor incentives. Simulated annealing is one of the technologies, through which the global optimum state is found in storage units, according to the local activation state of each neuron.
- **Self-organizing feature map:** The feature map is used to reduce the relay numbers of sensor signals and to save this topology in order to effectively integrate the sensor information from different sources.

Object-Oriented Programming

As with the above-mentioned forms of logical sensors, object-oriented programming is available to develop a unified framework by which the multi-sensor mission is completed. In many object-oriented applications of multi-sensor, each sensor represents a target. An object-based multi-sensor target identification system was presented by Gustafson et al. (2005), and an object-oriented framework was proposed which was used to

multi-sensor robotic missions (Delafosse M.& Clerentin A. et al., 2004).

Control Structure

Several different structures which are used to control the entire process of integration and fusion are introduced. Although these control structures have relationships with a particular application, they are also used in other areas.

Bayesian Network and Rule-Based System

Network and the rule-based system are the most commonly used forms of control structures for multi-sensor integration. They are used separately, and can also be used together. When the sensors in the system are different and multi-sensor data fusion needs to be done at multi-level, this structure is particularly effective. The rule-based system is the most effective in top-level control, while the network (such as Bayesian network or neural network) is the most effective in the bottom-level control. 'Decision-making network' was advocated to achieve multi-sensor integration (Huaizhi et al., 2004). In the decision-making network, the Bayesian network and neural network are used to judge the node of the tree structure.

The hierarchical structure is expressed effectively by using the Internet, such as the multi-assuming target identification; the layered network is not only used for objective modeling, but also is used to control the decision-making process. Many Artificial Intelligence-based control programs are achieved by using the rule-based system, so that multi-sensor integration of complex systems has high flexibility, and the production rule used in many systems is used for sign-level integration.

There are still problems in applications of rules-based systems in multi-sensor integration control structure at all levels, because once each rules in system and other rules are not independent, the reasoning process reaches an incorrect conclusion. The Bayesian formula is used to solve these problems because conditional probability is used to express true information or experience information for the Bayesian formula. To use conditional probability directly to express an existence problem, that is, before affirming a fact, the relevant conditions between this fact and other facts that have been known are required.

Multi-sensor target identification by the Bayesian network has been studied (Ronald & Yager, 2004; Gilbert, 2002). A continuous value estimation of the target state which is the output of multi-sensor (such as location and direction) is expressed in the form of a layered Kalman Filter in the network nodes.

NBS Layered Sensing and Control Structure

The Automatically Manufacturing Research Laboratory (AMRL) of the manufacturing engineering centre of the National Bureau of Standards in the United States has developed an interactive multi-sensor layered robot control system. The control system in the AMRL is composed by increasing the hierarchical 'sensing treatment' structure and dropping the 'task decomposition' layered control structure. A multi-layer structure is chosen because the complexity of the control procedure grows exponentially when the number of sensors and related processing increases. If some related handling parts are isolated at a certain level, the complexity is reduced. Therefore, it is possible to divide a large number of low-level processes (usually needed to do in real time) and a small number of high-level processes, so that the processing time needed by every floor is basically the same. Assuming the inter-layer communications are far less than that of the intra-layer, the complexity of the system will be reduced, because only a small number of communication channels are needed between the layers. If the processing of each layer is carried out parallel, the rise of layer numbers will not increase the complexity of the control procedures exponentially. Moreover, through the use of prior knowledge provided by the true model, the processing workload of each layer is greatly reduced.

Distributed Blackboard Structure

In the integrated multi-sensor system, the blackboard structure enables the distributed sensor subsystem to communicate economically. The time-sign cumulating is outputted to the blackboard by each subsystem to be employed by the fusion process and various integration functions. In the blackboard, the time-sign of the output makes the sensor information to be unified and dimensionless before fusion. All the system information needed by the integration function is supplied on the blackboard. A method of multi-sensor fusion was studied by Yenilmez & Temeltas (1998). It was used for the detection and identification of mobile robots. An object-oriented blackboard structure was used for reasoning (integration) and controlling and a solution comparing different methods of multi-sensor fusion by using the blackboard structure was proposed by Harris & Doyle (1997).

Because most of the tested objects have non-electrical power of different characteristics, such as temperature, pressure, sound, color and gray and so on, they are first converted into electrical signals and then transformed into digital signs that the computer is capable of handling through the A/D converter. The electrical signals require pre-treatment after being digitized, so that interference and noise in the data collection process are filtered. After treatment, feature extraction for a useful signal is needed.

Application of Sensor Integration and Data Fusion

Since most of the early studies of multi-sensor fusion focused on specific methods of enhancing computing power and the effective combination of data (information), and these studies were mainly based on military applications, so the technology was in a state of obturation for a long time. As the field of study and application expanded, the contents and results of related studies gradually appeared in all kinds of academic meetings and public documents, such as the United States armed forces data fusion annual, SPIE sensor fusion annual, the International Journal of Robotics and Automation, related IEEE meetings and journals and so on, making researchers of all areas increasingly aware of the importance of multi-sensor fusion technology. To date, multi-sensor fusion technology has been successfully applied to many manufacturing research areas.

Industrial Robots

In order to make industrial robots more sensitive, it is necessary to use several sensors. Most multi-sensor integration and fusion technologies are suitable for industrial applications.

The applications for industrial robots are generally divided into four areas: material handling, parts manufacturing (such as spot welding, arc welding, and forging), testing and assembly. Material handling is the simplest, and assembly is the most complex.

Material Handling

Industrial robots are utilized in industrial handling, to guide industrial trucks in the process of manufacture (such as an automatic guided vehicle or trailer) and the transmission of material handling equipment. Path planning and near-barrier capability of industrial trucks and material handling equipment is improved by using multi-sensor integration and fusion technology. Many multi-sensor integration and fusion methods which are used for mobile robots are employed directly in industrial handling trucks. A simulation system was designed to test the performance of a sensing control system in an automated material handling car (Gue & Kang, 2001). A robot has been constructed by which three-dimensional visual and force sensors are developed (Richards, 2005). It can start randomly placed connectors and put them on a printed circuit board. There is a research

project in the European Strategic Program of Research and development in Information Technology (ESPRIT) for a real-time material handling system with visual and tactile sensors.

Parts Manufacturing

According to 1985 statistics, approximately half of the robots in industrial areas of the United States are welding robots. Of these, spot-welding robots are in the majority, because sensing ability is needed in order to track a random changed weld line for arc-welding robots. The robot which is used has vision sensors and laser scanning distance sensors installed on the wrists, and the robot arms are guided by two kinds of sensors to weld along the welding seam between work pieces which feature random changes (Di X et al., 2004). Eddy current sensors and ultrasonic sensors are used in the development of the robots (Johansson & Robertsson, 2003) and a kind of sensor-guided arc welding system solution has been developed (Bauchspiess, A.& Absi A., et al., 1997). This system not only has the basic functions of robots (such as perception, place and crawl), but also has some advanced features of robots (such as identification and location).

Testing

Testing can be divided into two categories: direct testing and indirect testing. Direct testing mainly examines the integrity of parts as a separate process, is completed in the manufacturing process, and is accomplished after the completion of manufacturing. Various multi-sensor integration and fusion technologies introduced in this chapter all have the possibility of being used for direct testing, but which technology is used is determined in accordance with the work pieces. Indirect testing examines the integrity of the work pieces in the process of manufacturing. There are many methods of target identification, such as the use of visual and tactile sensors, and they are particularly effective in indirect testing. In the Georgia Tech Manufacturing Research Center, researchers are using visual and tactile information to realize the adaptive control of robot implementers. The technology is very effective in indirect testing.

Assembly

Assembly is the most complex aspect in robot applications. An assembly station has been introduced by Wang Y. H. & Jiang (2008) composed of two robots and a multi-function work piece feeding device. Each robot's wrist is equipped with visual sensors and force sensors, and a head camera. The head camera control robot assembly and the force sensors and vision sensors on the wrists are used to determine whether the installation of various components in place.

A multi-sensor robotic assembly station was introduced (Thomas *et al.*, 2007) which is equipped with visual sensors, ultrasonic sensors, touch sensors and force/torque sensors. The system is used for assembly of three different water pressure improvement parts of a heater. In the assembly process, visual sensors are used to distinguish various accessories, and feedback information from the force/torque sensors and feedback information from the robot crawling are used to control insertion and retention of the screw. After assembly, the visual sensors are used to test.

A multi-sensor robot workstation experimental system of graded control was presented by Thomas et al. (2004). It was developed by using the Rockwell automatic scientific test bed. The manipulation device at the extremity of the robot is equipped with visual sensors, acoustic sensors and force/ torque sensors. The limitation of information provided by each sensor and the superiority of information integrated by multi-sensors are proven through integration of different single sensors.

Intelligent Manufacturing System

The basis of an intelligent manufacturing system is intelligent machinery, which includes all kinds of intelligent machines, material transmitted tools and devices, detecting and testing equipment, and assembly devices. The purposes of intelligent manufacturing systems are to replace the mental work by using intelligent machines in manufacturing systems; to make the mental work be autonomous; to replace the operating skills of skilled workers with intelligent machines, so that the manufacturing process is no longer dependent on human 'skill', nor on surveillance and decision-making controls in the maintenance of automated production, so that the manufacturing system can be a self-production unit.

Intelligent Processing
Intelligent processing not only requires that the processing system has a high level of flexibility to maintain high productivity and the desired accuracy, but also that the control system is able to adapt to various processing conditions to achieve the most efficient or optimal process control. The manufacturing process is very complicated, however. It is usually a high-level non-linear system of multi-factors. Process models are often difficult to express in the form of an accurate mathematical model using analytical methods. Even if it is expressed as an analytical mathematical model, the robustness of the model is limited because the model is too complicated or because the conditions and assumed limitations are excessive.

State Monitoring
Condition monitoring system is an inseparable component of the intelligent manufacturing system. It effectively guarantees that the intelligent manufacturing system is operating normally by, for example, monitoring the wear of tools by using the multi-sensor integration of neural network (Kittler, 2001).

DATA FUSION

Classification of Systems

The information fusion structure is divided into different types, according to the information flow, control relationship, tactical application and integration scale (Lawrence A. Klein., 1993).

According to the Information Flow

From the relationship between the sensors and information flow of the fusion center, the information fusion structure is divided into four types:

Parallel Multi-Sensor Data Fusion
Parallel multi-sensor data fusion occurs when all sensor data are inputted to the same information fusion center. Sensor information can be current or from different times. Sensors can be similar in kind or different; information is from different times at the same sensor or from the same level and different levels at the same moment of the same sensor. The parallel type is more appropriate to solve the problem of multi-sensor data fusion of time and space, but there is a clear drawback, which is that the requested processing speed of the information fusion center should be quick, when the amount of inputted information is large. The information from the levels can be information from the same level and it can also be information from different levels. Most are information integration problems of the same level, because even if the inputted information is at different levels, they are converted into the same level to do the information integration.

Series Multi-Sensor Data Fusion
Series multi-sensor data fusion means that the information from two sensors is first integrated, and then the integration result and other sensor information are integrated. This operation is continued until all the sensor information integrations are accomplished. Forms of sensor information,

temporal and spatial parameters are the same with the parallel type. The current research on series multi-sensor fusion is small, because it is largely a multi-level pattern of two sensor parallel fusion, which obtains final conclusions by parallel type.

Hybrid Multi-Sensor Data Fusion

Hybrid multi-sensor data fusion is the combination of both series and parallel type. In its structure, parallel is behind series or series is behind parallel. It is similar to the mixed circuit composed of the parallel circuit and the series circuit. Inputted information is also the same as parallel. It has a variety of forms, and the operator is obtained by the combination of parallel type and series type.

Network-Based Multi-Sensor Data Fusion

The structure of network-based sensor data fusion is more complicated. From the relationship between the sensors and information from the sub-information fusion center, it is different from the above three types, it takes each of its sub-information fusion center as a node of the network. The input of this node contains the output information of other nodes, and may also contain sensor information flow. The final output is the output of an information fusion center or the output of a number of information fusion centers. The final conclusion is the combination of all these inputs. Sensor information is information of various levels or information of the same level. The input of the middle information fusion center is information at all levels.

The network-based type is divided into multi-storey type, horizon type and mixed type. Multi-storey multi-sensor data fusion is composed of an information fusion center of several layers, and each layer has a number of information fusion centers. A one-layer, two-layer or multi-layer information fusion center can be inserted between input and output, and these inserted layers are known as hidden layers. The input layer is usually composed of a number of information fusion centers, but it may also be formed from a fusion center, such as processing of multiple measurement information from a sensor. The information fusion center of each layer in the multi-storey type has no information relationship, but there are at least two existing information relationships among information fusion centers of each layer in the horizon type, where every information fusion center is a local estimator, completing a local estimate on the target. The mixed type is a combination of these various structures and is usually used in large-scale information integration systems.

According to Control Relationship

The control relationship of information integration is divided into open-loop and closed-loop.

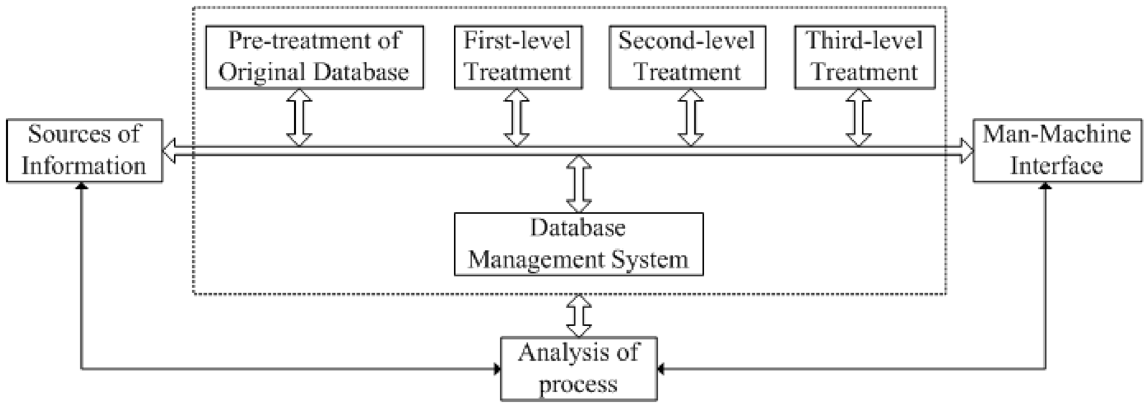
Open-loop multi-sensor information fusion is fusion in which there is no control on its sensors to work or on the information fusion center to fuse. The sensors collect, analyze and process information, in accordance with their own guidelines. The information fusion center treats sensor information comprehensively according to the existing integration rules, These processes are not subject to influences of the final conclusion or conclusions among the control, such as the various multi-sensor information fusion discussed above.

Closed-loop multi-sensor data fusion implies that treatment methods and judge rules of sensors or information fusion center are controlled or influenced by the final conclusions or central conclusions of the information fusion center, that is, the information processing has a feedback control process. Such feedback is either positive feedback or negative feedback. Feedback control can be divided into the sensor control and the control of the information fusion center.

Hierarchy of Data Fusion

Multi-sensor data fusion has three levels: the first layer is the location/status estimation, which is used to deal with all kinds of numerical data.

Figure 2. Processing model of multi-sensor data fusion



Location estimation is generally considered with optimal estimation technology (such as the Kalman filter) for the foundation, and status estimation is generally considered with parameter matching technology for the foundation, from the relatively simple technology (such as majority voting method) to more complex statistical methods (such as Bayes, or evidence reasoning). The second layer is situation estimation; the third layer is threat estimation. A large amount of abstract data (such as symbols) which reflects the relationship and meaning among numerical data is handled in the second and third layer, so using the symbol processing function of inference or reasoning technology is helpful to data fusion system technology.

Architecture of System

Data fusion technology has been widely applied to different areas of research and development. There are consequently different terms and definitions in different areas. In order to unify these terms, the U.S. military set up the Joint Directors of Laboratories (JDL) data fusion research group in 1986. The team developed the dictionary of data integration, and designed a model of the data integration process. Although this model has significant limitations, which need to be developed further, it has had an important impact on the basic understanding of data fusion. The model has been developed for the purpose of military applications, and it is now also used for related applications.

The basic components of the model are shown in Figure 2. The model shows the treatments of the process, and the types of functions and data types which are included in the entire integration process. In the model, the types of functions and the treatments of the process should not be separated; they should be integrated to form the entire integration function. The framework is introduced briefly below:

1. **Sources of information:** the information includes the sensors located in the scene of integration or placed distributed; it also includes prior information obtained from person or database.
2. **Pre-treatment of original database:** data show and data combination of different integration processes are implemented in this stage. Its purpose is to prevent the overload of inputted data in the integration processor, and the appropriate data is provided regularly

to the integration processor through the pre-treatment.

3. **First-level treatment (target refinement):** in this stage, the location, speed and entities of identification (such as military targets) are integrated. The stage includes four basic functions: data arrangement (data is converted by using a common unit of coordinates and units), tracing (analysis and location, velocity and other target information), the relevance of data (relevance of data and target) and identification (analysis of estimated identification of the target)
4. **Second-level treatment (form analysis):** the process of this stage draws conclusions through the use of relevant information, the links among targets, events, and prior information.
5. **Third-level treatment (trend analysis):** in this stage, inferences of future trends will be made, in accordance with the estimation of existing forms. The treatment is more difficult, because these inferences should not only be based on the results of actual calculations, but should also be based on the principles, tactics, strategy and environment of others.
6. **Analysis of process:** compared with other processes, the process is the secondary stage. The fusion process is controlled, through the surveillance and identification of integration implementation to improve information on system performance and reorganize sensors and resources in accordance with the goal tasks.
7. **Database management system:** this is a key component for a successful fusion system. The necessary features are: data retrieval, classification, compression, relation and data maintenance. The problem is difficult in some sense, because the management data is large and various, but retrieval must be quick.
8. **Man-Machine Interface:** this interface provides interactive tools. The human information obtained by the machines usually has the direct operation, information needs and evaluation results produced by integration systems. The information sent by machines to men includes warning, display of results and so on. The different natures of this exchanged information results in the demand for the ability of multimedia communications.

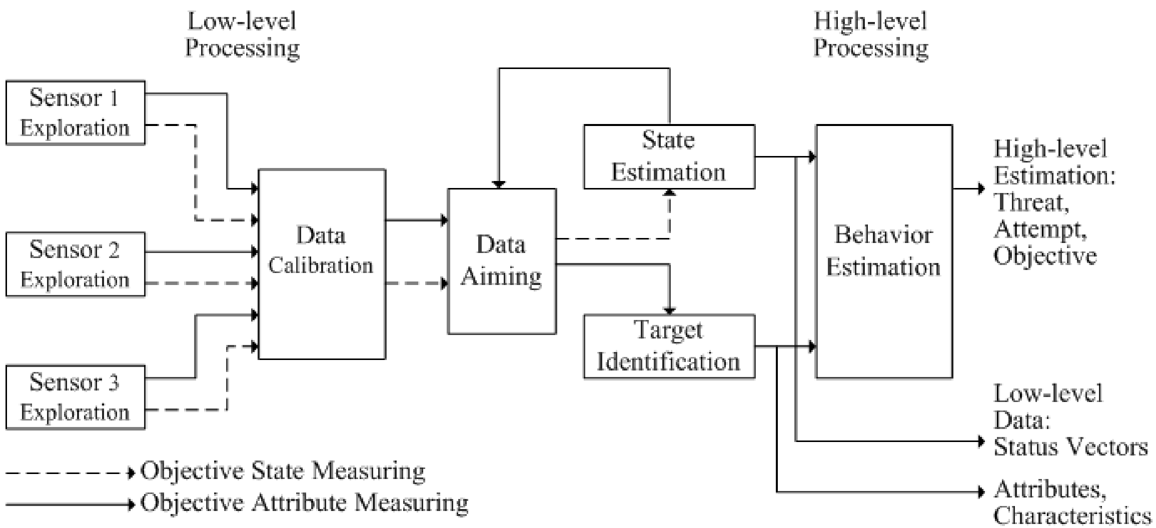
Every object of the JDL model can be further divided into different functional blocks and different mathematical models to complement their own. For example, the first-class treatment may include four function blocks - data combination, track, relation of data and identification - for each one; one or more mathematical models are used to achieve the above functions. For relation of data, doors, multi-channel track, the smallest neighbors or all the related technology will be used.

Function Model

In order to provide a general picture of the data integration process, a functional model of a simple multi-sensor data fusion system is presented in this section (shown in Figure 2), to illustrate the functional components of the universal integration system and the link between these functions (David L. Hall., 2001). The model assumes that there are only three sensors. They are monitoring the conjunct region which has more than one movement target of different types. For the sake of simplicity, only the basic information processing flow is discussed here, without involving intelligent processing and information management issues (such as databases and sensor management). The structures of the fusion system have various forms, but the role and theory of function are often the same.

In the model in Figure 3, the main functions of the data fusion system are calibration, correlation, identification and estimation. Calibration

Figure 3. Functional model of data fusion system



and correlation are prepared for identification and estimation, and the actual integration is conducted in the identification and estimation. The integration features of this model are completed in two steps, corresponding to different abstract information levels The first step is low-level processing, corresponding to pixel-level integration and feature-level integration; the output is the status, characteristics and attributes and so on. The second step is high-level disposal (behavior estimation), which are corresponded to the integration of the decision-making level; the output is the abstract result, such as threats, attempts, objectives and so on. The roles of the functional units will be introduced below.

Testing (Exploration)

The surveillance region is scanned by sensors. With each scan, all the objectives detected in the region are reported. Measurements and judgments of each sensor are independent. Once selected as a target, various measured parameters (objective catachrestic parameters and status parameters) are reported to the integration process.

Calibration (Aiming)

The role of data aiming units is to unify all the time and space reference points of sensors. If the sensors work independently and asynchronously in time and space, the calibration of time and space must first be carried out. The time transferring and coordinate transformation are employed to form the unified time and space reference points required by integration.

Relation (Internet, Association)

The role of data related units is to determine whether data of different times and spaces are from the same goal. At the end of each scan, the associated processing will be done among the collected new reports of some sensors and the new reports of other sensors and previous reports of the sensors by the relevant units. Using multi-sensor data to estimate the target, the data is required to be from the same goal. Such as using M_{jk} represents the number of goals, which is detected by the first j sensor at the moment of

k, then the data class of observation values of *s* sensors at the moment of *n* is:

$z+\{z_j\}$ $(j=1,2,\ldots,s)$; $z_j=\{z_{ji}(k)\}$ $(i=1,2,\ldots,m;k=1,2,\ldots,n)$ (1)

In the above formula, z_i means observation value collections of the first j sensor, $z_{ji}(k)$ means the observation value which is obtained from the first *i* target, by the first j sensor at the moment of k, m is the number of targets within the monitoring region.

State Estimation

State estimation is also known as target tracking. At the end of each scan, new data class and original data (obtained by the previous scan) will be integrated; the target parameters (such as location, speed) will be estimated according to observation values of the sensors, and by using these estimations, the target position of the next scan will be predicted. Forecast values are fed back to the next scan for related treatments. Output of the state estimation unit is the state estimation of objective, such as the state vector track.

Target Identification

Target identification is also known as attributes classification or identity estimation. An N-dimension feature vector is formed according to the goal characteristics measured by different sensors. Each one-dimension represents an independent feature of targets. If the goal has several types, and the features of each kind of target have been known in advance, then measured feature vectors and characteristics of known types will be compared to determine the types of targets. Target identification is considered as an estimation of objective attributes.

Behavior Estimation

The data classes of all the objectives (the target state and type) and behavior pattern of the possible situation previously established are compared to determine which behavior pattern matches best with the status of all the objectives within the monitoring region. Here, the behavior pattern is an abstract pattern, such as the attempt to target the enemy. It is divided into reconnaissance, attack, and so on. The output of behavior estimation units is situation estimation, threat estimation and trend, target attempt and so on.

Relation, identification and estimation processing functions exist through the entire integration system, which are the basic functions of the integration system. However, the orders using these functions have great influence on system architectures of the integration system, handling characteristics and performance.

THE METHODS OF DATA FUSION

The Weighted Average Method

The weighted average method is one of the simplest and most intuitive data fusion methods. Redundant information provided by several sensors after the weighted average is considered as fusion values. The dynamic values of original sensors are dealt with in real-time, but the workload of adjusting and setting weighted coefficients is very large, and is subjectivity to a certain degree.

Statistical Inference

Statistical inference is one of the simplest forms of uncertainty reasoning; the method is used for multi-source information integration, mainly for pattern recognition and classification of multi-source information.

Consistency Labeling Method

Consistency labeling is the establishment of corresponding relations between various objects and tags, when given a set of objects, a group of tags, a group of neighboring relations between the objects and a group of binding relationships between the tags, through an approach which is consistent with binding relationships. When the problem of multi-sensor data fusion is considered as a problem of consistency labeling, objects of labeling problems correspond to the characteristics of the goals measured by the sensors, and tags correspond to the objective description; at the same time, existing with the corresponding relationships between the tags and characteristics, there is also incredibility produced in the process of information acquisition and feature extraction, while the binding relationships between the tags are expressed by credibility.

Statistical Theory Based on Bayes

When the methods based on the statistical theory of Bayes are used for multi-sensor data fusion, the sensors are considered as different Bayes estimators, and they compose a decision-making system which has a symmetrical structure. Decision-making rules are used to select the best assumption estimation of the observed object.

Dempster-Shafer Evidence Synthesis Theory

Dempster-Shafer synthesis theory (evidence theory) as the expanding of Bayesian inference has been widely used in multi-sensor data fusion.

In Dempster-Shafer evidence synthesis, the basic entity is the identifying framework, and each information source is equivalent to an evidence body. Multi-sensor data fusion is essentially a process in which different evidence bodies are merged into a new evidence body in the same identifying framework. This merger is achieved by Dempster merger rules. Credibility is permitted to give the direct admission of evidence, whereas in the Bayesian approach, probability values are compulsority given to all evidences. This avoids simplifying assumptions made to unknown probabilities, while retaining the information and thus overcoming the defect of instability probability of the Bayesian method. In addition, an evidence interval is used in the Dempster-Shafer Evidence Rule (DSER), which expresses the information that is known and the uncertainties of the information, but in Bayesian, the knowledge and certainty of the information is explained. Therefore, DSER is widely applicable in multi-sensor data fusion.

Intelligent Fusion Method

In the process of multi-sensor data fusion, a large amount of abstract data (such as symbols) can be processed, which reflects the relationship and meaning among data, so synthesis and synthesis technology will be used. The symbol processing function of AI and expert system (ES) technology is used to obtain symbol synthesis abilities. ES owns information of various examples, and this information can be used to support traditional classification methods for status estimation. In multi-sensor data fusion, the key to the ES method is the engineering treatment of knowledge. Artificial intelligence technology has many important applications in multi-sensor data fusion, but because multi-sensor data fusion and AI technology are relatively immature, so there are still many of the issues to be resolved to make these two disciplines combine well. For example: time-varying dynamic input data and real-time operating requirements; various data types and knowledge types; processing and information transmission delay; spatial distribution of sensors; description of the authenticity of the background and expert experience; access and representation of knowledge; multi-level abstract of decision-making process; search technology, the size of knowledge base is too big, and so on.

APPLICATIONS OF MULTI-SENSOR INFORMATION FUSION

Cutting Tool Condition Detection

In the Cutting Tool Condition Detection system, as the source of the system information, the performances of the sensors have a direct impact on the success rate of the system detection. At present, through a lot of process detection research, people found that there are some detection methods which are widely-used and is valuable to promote, such as acoustic emission, cutting force, torque, vibration and machine power, and so on.

In the application of the methods above, the adopted detection strategy are divided into two parts: the model-based method and the character-based method. For the process, the sensor signals can be seen as the output of the dynamic system, therefore, we can detect the process through the establishment of the system model.

Three development directions were proposed (Dimia D.E. & Jr., 1997):

- Multi-sensor and multi-model system, it increases the information density of systems through the combine of multi-sensor and multi-model, in order to achieve a comprehensive and reliable detection of the process.
- Self-learning control system, it makes the detection system has strong nature to adapt to the changes in processing conditions, thereby enhance the system capabilities of self-adaptive and anti-jamming.
- Intelligent processing system, it can make decision according to the information of system state, and adapt to changes in the processing conditions well. In addition, it get information by the use of multi-sensor system; thereby can greatly improve the quality of decision-making.

Cutting Tool Condition Detection through Fuzzy Sets Theory

In order to improve the automation and intelligent level of processing system, we must develop effective and robust cutting tool condition detection system, to avoid the damage of machine tools, cutting tools and work piece. But in the process, because of the dynamic characteristics of machine tools, the cutting tool material conditions and the cutting conditions, the processing system will be highly non-linear and uncertain, and it is difficult to make an accurate mathematical model of the process. It is therefore difficult to get satisfying results by using traditional methods, and the focus has been changed to the intelligence of the process and the research of multi-sensor detection systems. The methods we have used include the method of pattern recognition, the statistical method, the method of artificial intelligence and the method of neural network. In the condition of fixed-cutting, the method of linear resolution function classification of pattern recognition can only be used in certain cutting conditions, because its method of signal classification is not universal. The artificial intelligence technology is capable of dealing with the uncertainty, but because of the slow response of the rules-based expert system, it is difficult to meet immediate requirements. Fuzzy detection technology provides a new solution to this problem.

Fuzzy set theory to detect the cutting tool condition is mainly chosen for the following reasons:

1. The machining process is a complex dynamic process and has more non-linear random interference, so the certainty relationship between the processing state and the detection state information which is established by using the traditional method is only applied to certain objects; it is widely used but its reliability is poor. By contrast, the detection model established by using the fuzzy method can avoid the restriction of

the detected objects, and is also suitable for dealing with the non-linear model, with better robustness.

2. Tool condition has a strongly ambiguous character as there are no clear boundaries to the early wear of the tool, normal cutting, medium-term wear and a serious wear state, so it is appropriate to identify them by using the soft classification method.

Compared with traditional detection technology, fuzzy detection technology has the following characteristics:

1. It uses ambiguous language variables to describe the expert's experience and knowledge, establish corresponding fuzzy rules and relationships, and to make these the basis for fuzzy reasoning and decision-making it is similar to human fuzzy reasoning abilities.
2. It identifies the state according to whether the membership function has good non-certainty and is suitable for state recognition in the non-linear system.
3. The classification and state identification is made by the simulation of human reasoning, which is based on expert knowledge modeling and is not dependent on the specific object model; thereby it does not need to establish a precise mathematical model of the detection objects, and has good adaptation. It is easy to implement in conditions where the expert knowledge and rules are complete, and also has a good reliability and robustness.

Ambiguity (also called relations function) is used to describe an uncertain incident in fuzzy set theory. That is, if *U* is for the domain, fuzzy set *A* is for the subset of *U*, then *A* is described as follow:

$$A=\{x|\mu_A(x)\} \tag{2}$$

$x \in U$ is for the value of A, $\mu_A(x)$ is for the value of ambiguity, also known as relations function, $\mu_A(x)$ is monotone function, $0\leq \mu_A(x)\leq 1$, as the A degree of uncertainty reduced, the value of $\mu_A(x)$ increase . If B is for a subset of U, and its ambiguity is higher than A, so:

$$\mu_A(x)\geq \mu_B(x) \tag{3}$$

If we set process state H for the input, and the monitoring indicators x for the output, the relationship between input and output can be described as the fuzzy linear equation below:

$$R=Q\circ p \tag{4}$$

r is for the ambiguity of monitoring indicators (x), p is for the ambiguity of cutting tool state (h), Q is for the fuzzy relationship function, "∘"is for Fuzzy operator.

Assuming the process has N different state (that is $h=(h_1,h_2,\ldots h_n)$ m monitoring indicators (that is, $x=\{x(1),x(2),\ldots x(m)\},h(x)=(h_1,h_2,\ldots h_n)$, that r is for an m-dimensional vector, p is for a n-dimensional vector, Q is for the $m\times n$-dimensional matrix which describes the fuzzy relationship between the status and indicators.

In another decision-making method, this method can also be divided into two parts: sub-study and classification. Study is the establishment of fuzzy relationship based on study samples (that is, Q is determined by p and r), and the classification is the vague of linear equations (that is to determine Q according to the known r and p).

Cutting Tool Condition Detection Based on Neural Network

To improve the system ability of information integration, optimizer algorithms are often used, such as the least-squares method and the greatest likelihood method. Through extensive research, it has been found that when using a signal algorithm to make the data registration and integration from

different information sources, it is hard to achieve the desired results. This applies to the cutting process in particular, which is of high uncertainty and non-linear; its information may be real-time, or may be non-real-time; it may be slow-changing, and may be transient; the relationship between the information may be interrelated, and may also be independent. The traditional information-processing method has thus been unable to adapt to information integration and integration needs based on multi-sensor, and has seriously hindered the development of intelligent detection technology. Neural network technology provides new ways in the development and research of the intelligent monitoring process.

Neural network using the Back Propagation (BP) network to detect tool wear was proposed to classify the new knife (0~0.25 mm) and the wear knife (0.5~0.75 mm) in the cutting process. Based on simulation, a system using neural network as the decision-making part in intelligent tool condition detection was proposed. In recent years, some researchers have continued research in this area, and have provided a basis for the research of intelligent detection technology in the process. Now, in order to explore the realization problem of neural networks in cutting tool progressive wear detection, neural network is utilized to achieve multi-sensor fusion of tool wear in the cutting process.

Because of the self-learning ability and pattern classification ability, the neural network is used progressively in the diagnosis of problems, such as the detection and diagnosis of tool wear and BP network process in particular is used most widely. Despite many examples of the successful application of BP, it still has some limitations in problem diagnosis in the manufacturing environment. Before being used, the BP network generally needs to be trained by a sample of different patterns, and if the network lacks prior knowledge of a failure state, it will be difficult for the network to detect this failure state. To deal with these problems, a variety of solutions have been proposed, such as the Adaptive Resonance Theory (ART) network without supervision learning, in which such problems do not exist. Another available network is the Radial Basis Function (RBF) network. Its hidden nodes use the RBF function as an incentive function. RBF function is normally a non-monotone function; that is, the hidden units make a meaningful response only when the input falls on a small designated space in the importation space.

Machine Thermal Error Compensation

Machine thermal distortion impacts on the accuracy of processing, in order to avoid too much inaccuracy in processing. There are many methods, such as the use of control equipment for temperature precision, materials of better thermal stability, and the production of machines of high-precision and good thermal stability by using advanced motors and bearings. In addition, quasi-steady geometric errors of machines are compensated for by the software, to improve the accuracy of machines. Quasi-steady geometric errors of machines are divided into two categories; one is caused by the static load, self-weights of the machine moving parts, manufacturing defects and inaccuracy of movements, and the other is deformed by machine thermal distortion. The former error can be solved by off-line error compensation of the known errors stored in the Computerized Numerical Control (CNC) controller in advance; this method has been widely used in industrial areas, such as the compensation of pitch values. Thermal error compensation has not been well-addressed in the industrial area. On one hand, this is because the existing CNC controller does not have the ability of real-time error compensation, and on the other it is due to the tedious and demanding work of obtaining a thermal error model, which is used in machines of on-line thermal error compensation.

In recent years, the Personal Computer (PC) has been used to make the real-time compensation for the thermal error compensation of machines. By using the PC to monitor the dynamic thermal errors of machines on-line, the compensation signals will be fed back in real-time to the machines, to edit the numerical control (NC) order. Generally, compensation technology is divided into direct compensation and indirect compensation. Direct compensation measures the offsets between the cutting tools and work pieces directly, and then makes compensation. As it is often not possible to measure the offsets in the process, direct compensation is usually difficult to achieve. Indirect compensation establishes a mathematical model through experimentation; the compensation is done by using the tool offsets which are calculated by this experience model. The experience model here is the relation equation between thermal errors and the temperature state of the machines, and thermal errors are obtained by the curve fitting of experimental data.

The difficulty of the thermal error compensation of machines thereby lies in the fact that the thermal error of the main edge of cutting tools is usually a comprehensive result, which is formed by thermal distortion of various components of machines that have different thermal properties, and the relationship between the processing status and thermal errors is highly non-linear, which brings on the difficulties of modeling.

Experience Model

The process is a dynamic process according to the above analysis, so it is necessary to estimate on-line the thermal errors of the process. A simple and effective way is to pre-build an experience model between the thermal error and machine temperature. In the process, the thermal error of machines will be estimated on-line by detecting the machine temperature.

Neural Network Model

In the process, thermal error is a dynamics, time-varying course which is highly non-linear. The model of its on-line prediction according to experience must therefore be a non-linear model with multiple input variables. Generally the model is established by using multi-variant regression analysis, and forecast accuracy is usually 60 percent. However, irrelevant and highly irrelevant variables of the multi-variant regression analysis model must be filtered out; otherwise, serious calculating issues will result in estimating model factures. Through the application of regression analysis techniques, this problem can be solved by automatic filtering of irrelevant and highly irrelevant items of the multi-variable regression analysis model, but it still is difficult for the modeling of multi-output variables.

The application of neural network technology is employed to build multi-input and multi-output relations between thermal error and temperature measurement values. The ability of the multi-output of the neural network not only greatly reduces the workload of the modeling, but also improves the accuracy of the model. Taking a three-layer feed forward network as an example, if hidden nodes are enough, any non-linear functions are capable of being fitted. Therefore, the feed forward network can be used to create a thermal error model, that is, the measured temperatures of machines are considered as the network input, and all kinds of thermal errors are considered as network output.

Since most CNC controllers are incapable of compensating for thermal error in real-time, the compensation values will be sent to the CNC controller in real-time with the help of an external PC. The following are several ways which are now commonly used.

Amend the Numerical Control (NC) Codes

The compensation principles of the method are introduced as follows.

Figure 4. Compensation principle of hardware amending method

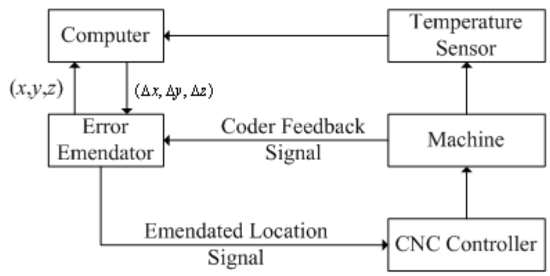
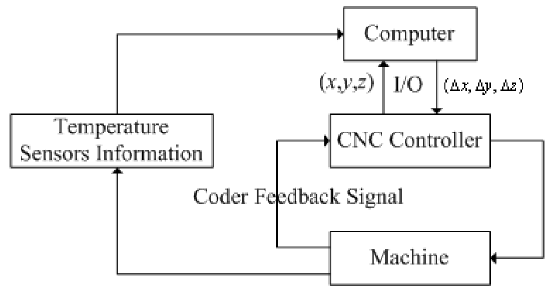


Figure 5. Compensation principle of realizing I/O communication



The NC program codes are firstly stored in the PC and order blocks are sent to the CNC controller by the PC using Distributed Numerical Control (DNC) technology. Before each order block is sent to the CNC controller, the current temperature state of machines is measured by computer to determine the appropriate thermal error according to the temperature state, and then by the use of thermal error, the NC order block is amended. The revised order block can then be sent to the CNC controller.

The disadvantage of this method is that it amends the endpoint error only, but high-speed machining is difficult to achieve because of the speed delay of the DNC.

Amend the Hardware

The compensation principles of the method are shown in Figure 4. The corresponding thermal error (Δx,Δy,Δz) is calculated in accordance with the temperature state of machines in the process, and then the error is sent to the error corrector; the position signals are then sent to the CNC controller by the error corrector to achieve the on-line compensation of thermal error.

The characteristic of this method is that in overcoming the existing problems in the process of amending the CNC codes, it can achieve real-time compensation; the shortcoming is the need to modify the hardware system, and for some of CNC machine tools, it is not applied.

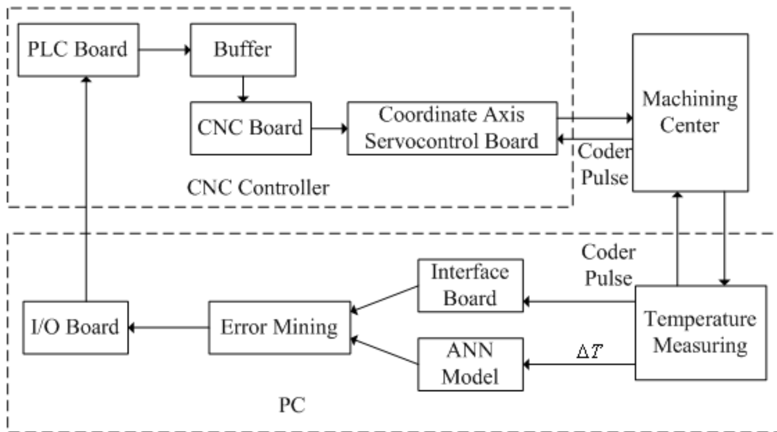
Real-Time I/O Communication

The compensation principles of this method are shown in Figure 5. The thermal error values are first calculated in accordance with the current temperature state of machines by computer, and then a compensation signal will be sent to the CNC controller in digital form through the I/O port, to achieve on-line compensation of the thermal error.

The characteristic of this method is that it can achieve real-time compensation without being required to transform the hardware of the servo system of machine tools, and is applicable to most of the current machine tools of CNC control.

The principle diagram of the thermal error compensation system is shown in Figure 6. Thermal error under the current temperature state of machine tools is forecasted, in accordance with a pre-established neural-network model, and the corresponding thermal error is then calculated in accordance with the position of slipped board. The compensation values will be given to the CNC controller through the I/O port, and these compensation values will be sent continuously to the coordinate reference registers of the coordinate axis by PLC (Programmable Logic Controller) control software. The coordinate origins of the coordinate's axis are adjusted in accordance with these compensation values by the CNC controller to achieve the compensation of thermal error.

Figure 6. Principle diagram of thermal error compensation system



Online Detection and Error Compensation of Ultra-Precision Sizing

Online detection technology of ultra-precision sizing is an important way to improve the efficiency and quality of ultra-precision sizing. It involves sensors and measuring instruments, measuring method and error description method, error separation and processing technologies, error evaluation technology and so on. Error compensation technology of ultra-precision sizing is a difficult procedure which involves the modeling of the process, the dynamic characteristics of ultra-precise implementation, ultra-precision positioning control and tracking control. This section focuses on the processing of cylindrical parts for ultra-precise gyroscopes cylindrical parts processing, which reflects a research achievement in which multi-sensor technology is used to achieve on-line measurement and error separation of cylindricity, spindle rotating error of multi-sensor ultra-precision machine tools, on-line measurement of the slide way line, and on-line multi-sensor measurement issues of the flatness of a large surface grinder.

Experimental Devices and Equipment

On-line measurement and error compensation of roundness and cylindricity are carried out on the S1-235 Chinese-made precision lathe. Using a hydrostatic spindle, the spindle rotating error of the lathe is about 0.5μm. An aluminum cylindrical work piece is cut using a diamond tool, and the surface roughness is up to Ra 0.016. Spindle rotating error depends on the manufacturing precision of the spindle, and the basic component of the spectrum is the component of 2 times the speed frequency, that is, the main error is ellipse error. 80 percent of the shape error of the work piece is spindle rotating error curve, but phase stagger is180.

In the S1-235 lathe, the work piece solid degrees are improved by the second-time processing error compensation which is done by piezoelectric ceramic tools. After the compensation, the resumption relations of error curves iss destroyed. On-line measurement sensors use high-precision capacitance sensors, and since the sensor is fixed, measured in the surface of movement tools, the information measured by the sensors is the sum of each instantaneous (sampling points) spindle rotating error and surface roundness error. The two errors are in the same order of magnitude, and it is

necessary to make three sensors distributed centripetally in the circular section. Through appropriate data processing, both errors can be separated, with accurate access to the spindle rotating error and roundness error of the work piece.

In a large high-precision surface grinder, online measurement and error separation of flatness degrees are realized by using multi-sensor fusion methods and the equipment of four sensors.

In the field of ultra-precision sizing, sensor fusion technology is expressed in the following two aspects:

1. Multi-sensor data fusion focuses on solving the above separation issues of work piece shape error and machine movement error.
2. A variety of error separation methods have their own shortcomings, so the characteristic-class integration of a number of ways can obtain more reasonable and more accurate results.

On-line Measurement Technology of Multi-Sensor Roundness and Cylindricity

In theory, to fully reflect all the information of cylindricity error, it is necessary to measure all the points of a cylindrical surface, but in practice this is impossible. Methods to obtain the error of a cylindrical surface have a cylindrical axis line (bus method) and cylindrical drive line. The bus method can be used in the cylindrical measurement with groove; the drive line method is divided into the section method and spiral method. Considering the methods and equipment by which the roundness error is obtained have been mature, the cross-section measurement of a number of equidistance is adopted.

The description of the Cylindricity of the Work Piece and the Mathematical Model

Any circular cross-section shape of work piece generated by rotating processing is expressed as the superposition of a base circle and error component. Error components show periodicity; it can be unfolded as the Fourier series, and it is shown as below:

$$R(\theta) = r_0 + \sum_{k=1}^{\infty} (a_k \cos k\theta + b_k \sin k\theta) \tag{5}$$

θ —polar phase

r_0 —the average radius of the work piece circular cross section

k —the numbers of harmonic waves

a_k, b_k—the Fourier coefficient of roundness error outline

Cylindricity On-Line Measuring Mathematical Model

The four-probe capacitive sensors we designed and developed are installed on the slipped board. In the process of measuring, the measuring shelf does the uniform-line movement, and the spindle folding work piece does the rotating movement. In addition to the error information of the work piece shape, the measured results also have two other errors, that is, the spindle rotating motion error and the carriage linear motion error caused by non-straightness of the slide way. In ideal circumstances, the impact on the work piece shape caused by machine movement error, in the curve unfolded as the time-axis, the latter is the former phase shift 180°. In the cutting of compensation control, the relations will be destroyed, so we need to accurately measure on-line the cylindricity of the work piece or the error caused by spindle rotation and non-straightness of the slide way, and error separation technology (EST) must be used.

The Cylindricity Error Assessment Method

The four methods of roundness evaluation are the least square method, the smallest region method, the largest in-circle method and the smallest circumcircle method. The cylindrical assessment method can also use all four methods. As conflict exists between the calculation efficiency

and accuracy of calculation in current calculation methods, many scholars have launched a variety of ways to improve the method, but have yet to reach a consensus.

According to shape and location tolerance national standards, the shape error is expressed by the width or diameter of the least tolerant region (or the smallest region for short). Accordingly, the roundness and cylindrical error are described as:

- **Roundness:** in the same section, the difference of the radius of two concentric circles, which contain the actual contours, and the difference of whose radius is the smallest.
- **Cylindricity:** the difference of the radius of two coaxial cylindrical surfaces, which contain the actual contours, and the difference of whose radius is the smallest.

Multi-Sensor Linear Error Measurement and Separation Technology

In the manufacturing error of the slide-way of the machine tool, the relative sliding between the splinted board and slide-way may have six degrees of freedom. When processing a cylindrical work piece in an ultra-precision diamond cutting lathe, the cutting dynamic model is a complex process, so the error through the tool movement is not completely copied to the work piece, and forming the generatrix of the work piece is difficult. This is because the imprecise movement of the splinted board depends not only on the geometric shapes of slide-ways, but also on cutting conditions, cutting load, thermal distortion and movement direction. The online measurement method of the work piece straight line is usually to install the sensors on a splinted board; the motion error of the slipped board will then be introduced to the measurement data, and the uncertainty elements of splinted board movement will have an impact on the measurement repeatability.

The Application of Multi-Sensor Fusion in Large Flatness On-Line Detection

In the large ultra-precise surface grinder, the multi-sensor fusion method is used to measure the surface errors of the large ultra-precise work piece. Typically, sensor groups are installed in the grinding moving parts to measure the surface of the work piece line by line; the data measured by each sensor include the flatness error of the work piece surface and the motion error of the slide-way. The multi-sensor fusion method can separate the above two errors, by which a three-dimensional 'map' of flatness error and motion error are drawn. These data can therefore be used to do the machining compensation, to achieve the integration of computer measurements and the machining technology. This method is also adopted for large-scale flatness measurement instruments.

CASE STUDY

In this section, we will present a case study about dynamic positioning method for parallel machine (PM) based on Kalman Filtering (KF) data fusion.

Background and Motivation

Parallel mechanism is a hot topic in the field of Robot Research, as well as in the field of the Machine Tool manufacture. The static platform and dynamic platform of PM are supported and connected by many poles (legs), spindle and tools are mounted on dynamic platform, and their poses (positions and attitudes) change correspondingly with the changes in the length of poles.

The key technology impacting spindle positioning accuracy of parallel machine is the technology measuring accurately the changes in the length of each leg. In general, changes in the leg's length of the PM can be measured accurately by using

laser interferometry. But it is difficult to be applied widely because of high cost. Another measuring way is using encoder. Because encoder can only measure screw rotary numbers for conversion, and cannot measure the deformations caused by the factors such as mechanical impact, abrasion and thermal expansion, leg length measurements are inaccurate, and the errors will be passed to the poses of dynamic platform by Forward Kinematics Equations, leading to the positioning accuracy decline. So the study of new measurement method has great significance for improving the accuracy of PM and promoting the practice of PM.

Inertial sensors such as accelerometers and gyroscopes have been widely used in the navigation system, to provide oriented information (position and direction) for the navigation system. In recent years, the inertial sensors have also been applied in manufacturing field, such as industrial robots and machine calibration. Inertial sensing technology can meet the precision requirements in the process of demarcating the deviations of movement directions such as tilt, swing and deflection. It is presented in this section about a new dynamic positioning method of PM based on KF data fusion. In this method, the poses of dynamic platform are measured directly by using inertial sensor, and the measured inertial sensor data and encoder data are fused by KF algorithm to compensate measurement and calculation errors, so that it can enhance the positioning accuracy of dynamic platforms.

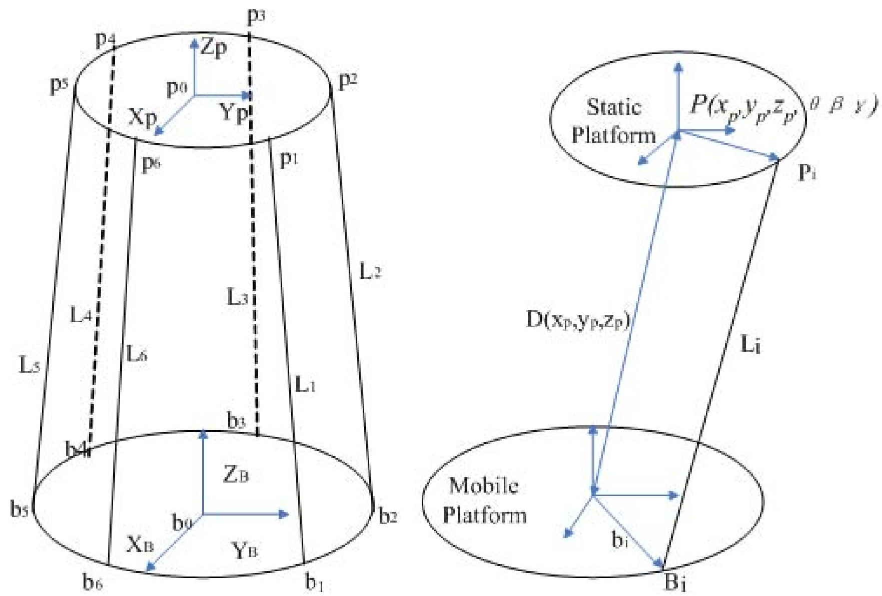
Dynamic Positioning Method for PM Based on KF Data Fusion

Dynamic Positioning Method about the Dynamic Platform of PM

In general, the kinematics of PM is analyzed and controlled through the positive solutions and inverse solutions of kinematics model. Since it is difficult to get the closed-loop solution for positive solutions of the kinematics model, but can only be expressed with high-order polynomial equation, which to a large extent cannot obtained the pose of dynamic platform in real time. To this end, it is presented in this section that the pose of dynamic platform obtaining spindle is measured by using inertial sensor technology. In this method, inertial measurement units containing three accelerometers and three gyroscopes are installed on dynamic platform in each vertical, shown as Figure 7. In Figure 7, B is the coordinate system of the static platform; P is the coordinate system of the dynamic platform. p_i and b_i present dynamic platform nodes and static platform nodes separately, leg length vectors of PM are L_i, i= 1~6. Tool centre pose (TCP) of the dynamic platform (x_p, y_p, z_p, θ,β,γ) is obtained by integrating the dynamic measurement acceleration and angular acceleration. On the other hand, by using dynamic inertial system to measure the pose of dynamic platform directly, spindle TCP of the PM can be obtained by measurement, not needing the Forward Kinematics. The TCP of PM is measured by the inertial sensors, and its location and direction are obtained through the numerical points, and then to assess and correct by KF, so that the positioning accuracy is further improved. Comparison of measured TCP and the set TCP, the differences feedback is returned to the controller and the length of each leg is adjusted by the inverse solution of kinematics equation. Therefore, it is considered that the direct measurement and obtaining access of TCP of the dynamic platform is an effective way to greatly simplify the positive solutions of kinematics equation.

As a result of the impact on manufacturing error, system error and environmental factors, inertial measurement system contains sensor error, sensitivity error, and the impact on noise and other random errors. The error included in the measurement data will be fused with the acceleration data to do the data processing, the error after numerical integration increases over time growth, and reduce the measurement data accuracy. Growth over time, even the tiny offset of acceleration can be a lot of

Figure 7. Installation of inertial measurement sensor in PM



position error. Therefore, the method measuring data error control and compensation is the key of TCP direct measurement of PM.

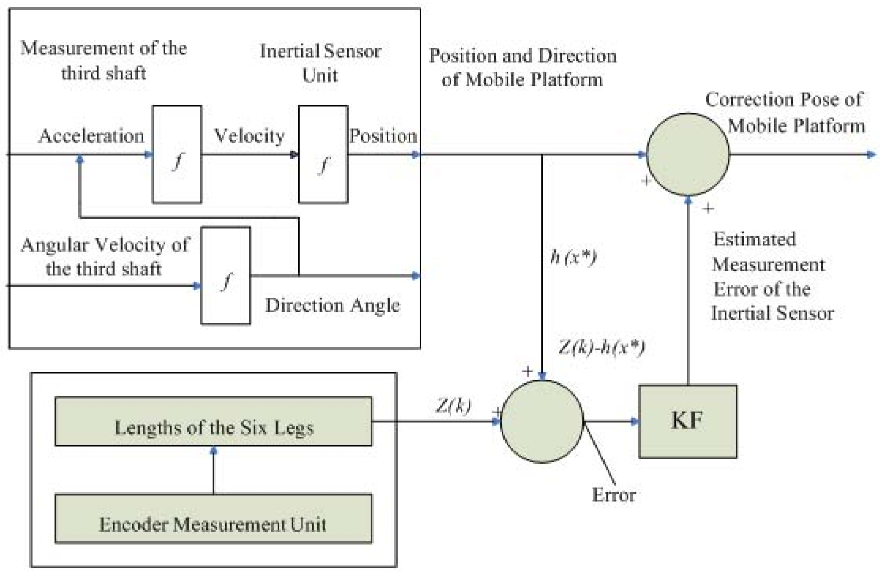
Dynamic Positioning Algorithm of Dynamic Platform Based on KF Data Fusion

The proposed dynamic measurement system contains inertial measurement data and measurement data of encoders in each leg of the machine. For the encoder data, only the length changes can be calculated through screw revolution, but no any dynamic information relevant to the length changes of screw measured, such as deformations caused by mechanical impact, abrasion, thermal expansion and so on. In contrast, inertial sensors can directly dynamic measure the pose changes of dynamic platform, to obtain any information about the pose changes of dynamic platform caused by screw deformation, but there will be error drift in its data process, so the accuracy is affected. Since these two types of data are fused, there could be made up, and the direct measurement accuracy of the dynamic platform pose can be improved. By using KF statistics technology to achieve the integration between inertial measurement data and external encoder data, the algorithm will output the optimal result of pose dynamic measurement.

Figure 8 presents the pose measurement system framework of the dynamic platform of PM based on the KF algorithm, it shows that the algorithm structure on the KF-based pose measurement system of the dynamic platform of PM. $z(k)$ is the measurement variable of leg length, $h(x^*)$ is the transformation matrix to map the system state variables into measure variables, the symbol $\int$ represents one integral. The establishments of inertial measurement system model and encoder measurement model, and the export of relational matrix between models are introduced separately as follows.

In the inertial measurement system of the dynamic platform of PM, the pose TCP of the dynamic platform is described by six degrees of freedom (DOF), therefore, six inertial sensors (three accelerometers and three gyroscopes) are needed to use to measure the pose changes of the

Figure 8. Framework of pose measurement system of PM dynamic platform based on KF algorithm



dynamic platform, and one accelerometer and one gyroscope in each direction are installed to measure the acceleration and angular velocity in the direction. To simplify the description of the problem, first PM with six DOF is simplified to express as the movement and rotation in a single direction, and then the model is extended to the machine with six DOF.

In KF algorithm, measurement model is usually expressed as Equation 6.

$$z(k)=h(x(k))+v(k) \quad (6)$$

z(k) - *m**1-measurement variables at the time *t(k)*;

h(.) - transformation matrix to map the system state variables into measure variables;

v(k) - *m**1-measurement noises relevant to measurement uncertainties, assuming the noise is an unrelated random sequence, its mean is 0, namely $E[v(k)]=0,\forall k$, and its covariance is:

$$E[v(k)v(i)^T] = \begin{cases} R(k) & i = k \\ 0 & i \neq k \end{cases} \quad (7)$$

And assuming the Gaussian noise series of the sensor measurement system and the Gaussian noise of the external encoder measurement are independent and unrelated.

$$E[w(k)v(k)^T]=0 \quad (8)$$

For all the *k* and *i*, measurement equation can be unfolded by Taylor coefficients, namely

$$z(k)=h(x(k))+v(k)=$$

$$h(\bar{x}(k)) + \frac{\partial h}{\partial x}\Big|_{x=\bar{x}} \delta x + \text{Quadratic term + high-order term} + v(k) \quad (9)$$

If neglecting quadratic term and high-order term, Equation 10 can be expressed as

$$z'(k) = z(k) - h(\bar{x}(k)) = H(\bar{x}(k))\delta x + v(k) \quad (10)$$

Here, $H(\bar{x}(k)) = \frac{\partial h}{\partial x}|_{x=\bar{x}}$ is measurement relational matrix, describing the theoretical relation between measurement variables and system state variables in time *t(k)* (excluding noise conditions).

In this system, the six leg lengths of the PM are measured by encoder, if L_i represents the lengths of the legs l_i, there are measured variables z=*L*= $(L_1\ L_2\ L_3\ L_4\ L_5\ L_6)^T$. However, since there are no direct links between the measurement model and system state variables, the relationship between these external measurement variables and system state variables must be set up, in order to achieve the data fusion of the inertial measurement data and encoder measurement data, and to achieve the purpose of dynamic measurement.

Relational matrix *H(k)* can be defined by matrix $\partial h / \partial x$, and the matrix $\partial h / \partial x$ is obtained through each leg length l_i of PM ordering partial derivatives against the pose variables respectively. Relational matrix *H(k)* can is defined in Equation 12.

$$H(k) = \frac{\partial h}{\partial x} = \begin{pmatrix} \frac{\partial h_1}{\partial x_p} & 0 & 0 & \frac{\partial h_1}{\partial y_p} & 0 & 0 & \frac{\partial h_1}{\partial z_p} & 0 & 0 & \frac{\partial h_1}{\partial \theta} & 0 & \frac{\partial h_1}{\partial \beta} & 0 & \frac{\partial h_1}{\partial \gamma} & 0 \\ \frac{\partial h_2}{\partial x_p} & 0 & 0 & \frac{\partial h_2}{\partial y_p} & 0 & 0 & \frac{\partial h_2}{\partial z_p} & 0 & 0 & \frac{\partial h_2}{\partial \theta} & 0 & \frac{\partial h_2}{\partial \beta} & 0 & \frac{\partial h_2}{\partial \gamma} & 0 \\ \frac{\partial h_3}{\partial x_p} & 0 & 0 & \frac{\partial h_3}{\partial y_p} & 0 & 0 & \frac{\partial h_3}{\partial z_p} & 0 & 0 & \frac{\partial h_3}{\partial \theta} & 0 & \frac{\partial h_3}{\partial \beta} & 0 & \frac{\partial h_3}{\partial \gamma} & 0 \\ \frac{\partial h_4}{\partial x_p} & 0 & 0 & \frac{\partial h_4}{\partial y_p} & 0 & 0 & \frac{\partial h_4}{\partial z_p} & 0 & 0 & \frac{\partial h_4}{\partial \theta} & 0 & \frac{\partial h_4}{\partial \beta} & 0 & \frac{\partial h_4}{\partial \gamma} & 0 \\ \frac{\partial h_5}{\partial x_p} & 0 & 0 & \frac{\partial h_5}{\partial y_p} & 0 & 0 & \frac{\partial h_5}{\partial z_p} & 0 & 0 & \frac{\partial h_5}{\partial \theta} & 0 & \frac{\partial h_5}{\partial \beta} & 0 & \frac{\partial h_5}{\partial \gamma} & 0 \\ \frac{\partial h_6}{\partial x_p} & 0 & 0 & \frac{\partial h_6}{\partial y_p} & 0 & 0 & \frac{\partial h_6}{\partial z_p} & 0 & 0 & \frac{\partial h_6}{\partial \theta} & 0 & \frac{\partial h_6}{\partial \beta} & 0 & \frac{\partial h_6}{\partial \gamma} & 0 \end{pmatrix} \quad (11)$$

Here, the partial derivative component against location is

$$\frac{\partial h_1}{\partial x_p} = \frac{\partial l_1}{\partial x_p} = \frac{x_p - b_{1x} + r_{11}p_{1x} + r_{12}p_{1y} + r_{13}p_{1z}}{\sqrt{l_1^2}} = \frac{L_{1x}}{l_1} \quad (12)$$

Similarly, the remaining partial derivative components are obtained

$$\frac{\partial h_1}{\partial y_p} = \frac{L_{1y}}{l_1} \quad (13)$$

After establishing the inertial system model and measurement system model, and confirming the transition matrix *H(k)*and *ϕ(k)*, the covariance matrix *Q* and *R* used to measure noise must be calculated, and the value *x(0)* of the initial state variable and initial covariance *P(0)* must be given, so as to achieve the iterated algorithm about KF. During the implementation process, it is necessary to adjust the covariance values *Q* and *R* to adapt to the measurement environment and track changes.

Conclusion

1. Inertial sensor technology is applied to pose measurement of dynamic platform of the PM, based on KF data fusion algorithm, a new method on pose direct measurement of the dynamic platform of the PM with six DOF is proposed. The method makes full use of the characteristic that inertial sensor technology can realize real-time and dynamic measurement. For the inherent error drift problem, it can be revised by using encoder measurement, and good positional accuracy of dynamic measurement can be made through data fusion, at the same time dynamic performance can be reflected also, propitious to the stability control of PM.
2. They are described in detail about the establishment process of the inertia system model and the external measurement model in the KF algorithm, as well as the derivation process of the relational matrix between measurement variables and state variables. The algorithm availability has been initially confirmed through a uniaxial movement test, and six DOF pose measurement of the dynamic platform has been simulated and researched.

3. The experiment and research have shown that KF-based data fusion algorithm can be used for the pose direct measurement of dynamic platform, to obtain the dynamic state properties, but the control of measurement error is to be further improved.

SUMMARY

Multi-sensor integration and fusion technology is actually a comprehensive technology of multi-source information; through the analysis and integration of data information from different sensors (information sources), the best consensus estimate of the measured target and its nature is achieved. The development state, technical background, application scope and basic meaning of the multi-sensor integration and the data fusion have been presented in this chapter. The classifications, layers, system structure and function model of the data fusion system have also been discussed, and the theoretical method of data fusion introduced. Finally, the application of multi-sensor data fusion technology in intelligent manufacturing was focused on.

Multi-sensor integration and fusion technology is a booming research domain. It is multi-disciplinary. It relates to signal processing, probability statistics, information theory, pattern recognition, fuzzy mathematics, and more. The basic requirements of the information fusion method are robustness and parallel processing because the information of the multi-sensor system is diversiform and complicated. Additionally, arithmetic speed and precision, interface performance, cooperation, and information sample are important.

Much more useful information is gained from the environment via a multi-sensor system. The development tendency of multi-sensor integration and data fusion are described as follows:

1. Establishing the basic theories of data fusion. This includes further research on the basic mathematics of fusion techniques and the numerical treatment of congener information fusion. Every optimum and suboptimum distributed scheme is researched. Some new technologies are developed to deal with the symbol processing method for inhomogeneous information fusion, such as new AI which are self-learning, evolution arithmetic, wavelet analysis and evolutional neural networks. The combination of evolution arithmetic and neural network has been used in time series analysis-based fiscal budget systems.
2. Using AI methods and knowledge-based multi-sensor fusion ways in data fusion technology are a future research trend, because AI is adept in intelligibility and in dealing with complex questions.
3. An important development tendency is more organic and effective integration among fuzzy mathematics, neural networks, evolutionary computing, rough set theory, wavelet transformation and expert systems.

REFERENCES

Bauchspiess, A., Absi, A., et al. (1997, September 7-11). Predictive Sensor Guided Robotic Manipulators in Automated Welding Cells[J]. In *IROS'97., Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Vol. 2, pp. 1144-1150).

Dasarathy, B. V. (1997). Sensor Fusion Potential Exploitation Innovative Architectures and Illustrative Applications [J]. *Proceedings of the IEEE*, *85*(1), 24–38. doi:10.1109/5.554206

Delafosse, M., Clerentin, A., et al. (2004, December 8-10). Multi Sensor Fusion and Constraint Propagation to Localize a Mobile Robot. *IEEE ICIT '04. 2004 IEEE International Conference on Industrial Technology*, (Vol.1, pp. 66-71).

Di, X., & Song, M. (2004). Realtime Integration of Sensing, Planning and Control in Robotic Work-cells [J]. *Control Eng. Practice, Istanbul, Turkey*, *12*(6), 653–663.

Dimia, D. E., Jr. (1997). A Multi-sensor Integration Method of Signals in a Metal Cutting Operation via Application of Multilayer Perceptron Neural Networks[J]. In *Fifth International Conference on Artificial Neural Networks, (Conf. Publ. No. 440)*, July 7-9, (pp. 306-311).

Figliolini, G., Ceccarelli, M., et al. (2000). Logical Sensors and Control System Programming for an Autonomous Biped Walking Robot. In *IROS 2000, Proceedings of 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 31-November 5, (Vol. 2, pp. 1346 -1351).

Gilbert, S. (2002). Data Fusion and Data Grafting [J]. *Computational Statistics & Data Analysis*, *38*, 465–473. doi:10.1016/S0167-9473(01)00072-X

Gue, K. R., & Kang, K. (2001). Staging Queues in Material Handling and Transportation Systems. In *Proceedings of the Winter Simulation Conference*, December 9-12, 2001, (Vol. 2, pp. 1104-1108).

Gustafson, E. H., Lollini, C. T., et al. (2005). Swarm Technology for Search and Rescue through Multi-sensor Multi-viewpoint Target Identification. In *SSST '05. Proceedings of the Thirty-Seventh Southeastern Symposium on System Theory*, March 20-22, (pp. 352-356).

Hall, D. L. (2001). *Handbook of Multisensor Data Fusion* (*Vol. 6*, pp. 85–110). Boca Raton, FL: CRC.

Harris, C. J. & Doyle, R.S. (1997, February 18).Multi-sensor Data Fusion for Situational Assessment—a Critical Element of Systems Integration, Some Theory and Application to Collision Avoidance. *IEEE Colloquium on Integrated Systems in Aerospace, 5*(1-5), 3.

Henderson, T., & Hansen, C. (1985). A Framework for Distributed Sensing and Control. In *Proceeding of 9th International Joint Conference on Artificial Intelligence, Los Angeles,* Aug. (pp. 1106-1109).

Henderson, T., & Shilcrat, E. (1984). Logical Sensor System. *Journal of Robotic Systems*, *1*(2), 169–193. doi:10.1002/rob.4620010206

Huaizhi, Y., Changzhen, H., et al. (2004). Multi-sensor Information Intelligence Fusion Model Using Fuzzy Colored Petri Nets. In *WCICA 2004. Fifth World Congress on Intelligent Control and Automation,* June 15-19, 2004, (Vol. 4, pp. 3072-3075).

Johansson, R., & Robertsson, A. (2003). Robotic Force Control Using Observer Based Strict Positive Real Impedance Control. In *IEEE Proc. Int. Conf. Robot. Autom.,* Taipei, Taiwan, R.O.C., (pp. 3686–3691).

Kittler, J. (2001). Multi-sensor Integration and Decision Level Fusion[J]. In *A DERA/IEE Workshop on Intelligent Sensor Processing (Ref. No. 2001/050),* February 14, (pp. 6/1-6/6).

Laconte, J., Flandre, D., et al. (2006). Micromachined Thin-Film Sensors for SOI-CMOS Co-Integration. New York: Springer US. Klein, L. A. (1993). Sensor and Data Fusion Concepts and Applications (pp. 90-100). Bellingham, WA: Society of Photo-Optical Instrumentation Engineers (SPIE).

Richards, J. A. (2005). Analysis of Remotely Sensed Data: The Formative Decades and The Future. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(3), 422–432. doi:10.1109/TGRS.2004.837326

Ronald, R. & Yager. (2004). A Framework for Multi-source Data Fusion [J]. *Information Sciences*, *163*, 175–200. doi:10.1016/j.ins.2003.03.018

Thomas, U., Florke, J., et al. (2004). An Integrative Approach for Multi-sensor Based Robot Task Programming[J]. In *Proceedings of ICRA '04. 2004 IEEE International Conference on Robotics and Automation,* April 26-May 1, (Vol. 2, pp. 1149-1154).

Thomas, U., Molkenstruck, S., et al. (2007). Multi Sensor Fusion in Robot Assembly Using Particle Filters[J]. *IEEE International Conference on Robotics and Automation,* April 10-14 (pp. 837-3843).

Wang, Y. H., Jiang, X. H., et al. (2008, April 6-8). Research on the Method of Intelligent Robot Visual Recognition and Positioning[J]. In *ICNSC 2008. IEEE International Conference on Networking, Sensing and Control*, (pp. 916-919).

Yenilmez, L., & Temeltas, H. (1998). *Real Time Multi-sensor Fusion and Navigation for Mobile Robots*. In *MELECON 98, 9th Mediterranean Electrotechnical Conference*, (Vol. 1, pp. 221-225), May 18-20.

Chapter 8
Group Technology

Group technology (GT) is a management philosophy that attempts to group products with similar design and/or manufacturing characteristics. It is also a key factor in the successful implementation of flexible manufacturing systems, and equally is one of the foundations of the implementation of intelligent manufacturing. The success of GT implementation is in the effective formation of part families and the rational layout of the manufacturing cell (machine family). In this chapter, the background and conception of (GT) are introduced, followed by succinct descriptions of the similarity criterion, classification and coding systems, and classification approaches of GT. The actual applications of GT to product design, process planning and group scheduling are discussed, and finally the summary and trends of GT are articulated.

DOI: 10.4018/978-1-60566-864-2.ch008

INTRODUCTION

The Background and Conception of GT

Mass production was considered to be a virtue in the 1960s and 1970s, but this situation changed during the 1980s. Multi-product, small-lot-sized production gradually became the preference from 1980s, and to date the mass customization of products has become an effective manufacturing mode for enterprises to seize the advantage in a strongly competitive environment. Applying the traditional means of production to multi-product, small-lot-size production and mass customization of products incurs many difficulties, such as complexities of organization and managment, difficulties of production planning and scheduling, as well as limitation of the application of advanced

technology. To cope with these difficulties, several effective approaches have been developed. Of these approaches, and in addition to flexible manufacturing and standardization, simplification, and specialization of production, GT plays an important role in multi-product, small-lot-sized production and mass customization of products. There have been many applications of GT in the process of production, from design and planning to the organization of manufacturing facilities. In practice, GT provides a link between design and manufacturing. Various countries demonstrated different interests in GT in the early research, for example, the focus in Russia was on the grouping of parts for processing on single machines, mainly to rationalize tooling and to reduce resetting times; in the UK and Germany, the greatest emphasis has been on the development of cellular layout and group flow line; in Japan, concentration was not only on developing classification systems and group layouts, but has also extended the basic approach to flexible manufacturing systems; in the U.S.A., emphasis has been placed on component analysis and associated data processing (Gallagher, 1986). However, what is the definition of GT? Shunk defines GT as a disciplined approach to identifying things such as parts, processes, equipment, tools, people, or customer needs by their attributes and looking for similarities between and among the things: grouping the things into families according to similarities; and, finally, increasing the efficiency and effectiveness of managing the things by taking advantage of the similarities (Shunk, 1985).

The application of GT results in acquiring the advantages of multi-product batch manufacturing in the small-lot-sized production and the mass customization of products. With the advent and development of CIMS (Computer Integrated Manufacturing System), GT become the bridge to span CAD (Computer-Aided Design), CAPP (Computer-Aided Process Planning), and CAM (Computer-Aided Manufacturing). GT has now become a kind of production technology science, which takes full advantages of similarities of parts in the process of production to increase production efficiency, *viz*., the similar approaches are used to fabricate the part families formed by grouping similar parts. The crux of GT is concerned with how to identify and explore the similarities, and to reduce the time and cost of manufacturing by the sharing solutions presented. GT is a concept for increasing production efficiency by grouping various parts and products with similar design and/or production process (Ham, 1985). It is also an approach to organizing production or a manufacturing philosophy. The philosophy of GT has been used to small-batch variable production. The criterion has been applied in many fields including machining, welded fabrications, foundry work, presswork, forging, and plastic injection moulding. The basic approach enables all aspects of manufacturing, from design, through estimating and planning to production, to be rationalized (Gallagher, 1986). The overall advantages have been summarized by Andrew Kusiak (Kusiak, 1990). The main advantages are as follows.

- Reduced production lead time (20-88%).
- Reduced work-in-process (up to 88%)
- Reduced labor (15-25%)
- Reduced tooling (20-30%)
- Reduced rework and scrap materials (15-75%)
- Reduced setup time (20-60%)
- Reduced order time delivery
- Improved human relations
- Reduced paper work

From the birth of GT to the present has been about 60 years. During this time, GT has been laying one of the foundations for the application of computer-aided technology in manufacturing. The application of GT in manufacturing changes the paradigms of management and the modes of processing.

Table 1. Various classification characteristics and applications

Part Families	Similarity Features	Applications
Designing families	Part functions, geometric shape, material type, shaping type, and machining characteristics.	Part drawing retrieve, rationalization and standardization of design, CAD, and etc.
Processing families	Geometric shape, process, material type, machining dimension, facilities, processing equipments.	Cell layout, group process planning, group jig design, CAM, and etc.
Managing families	Productive organization of group technology, batch, time of manufacturing command, complexity of manufacturing	Production management, adjustment of machine loading, planning arrangement, alterative blue print, computer-aided management system.

Similarity Criterion

The core of GT is the similarity criterion, namely, that there exist similarities in the process of design, process planning, and production management of parts, which seeks to utilize fully the similarities of geometric shape or process features of parts to organize production. Consequently the manufacturing costs and setting-up times are reduced, and the production efficiency is improved. In grouping, parts are classified into various part families based on different criteria, which becomes the technical foundation for the implementation of GT. The application or adoption of GT starts with identifying part families and machine groups so that each part family is processed within a machine group with minimum interaction with other groups. Part families are a collection of similar parts, and different part families are presented in terms of different similar characteristics. For example, parts are classified into design families, process families, and management families according to these characteristics. Table 1 shows the various classification characteristics and applications of different part families.

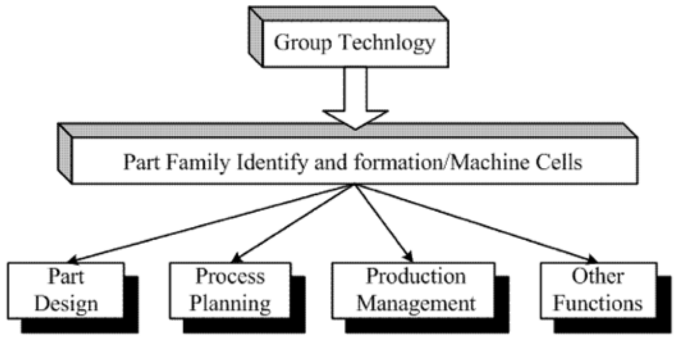
The philosophy of GT is to expand the production batch, namely, to transform multi-product, small-lot-sized production to large-lot-sized production using the similarity criterion. It has been reported that in batch-type metalworking shops, only some 5% of the total production time is spent on machine tools while the remaining 95% is spent in moving and waiting for parts in the shop. Of that 5% time, only about 30% is spent as productive time in cutting materials, and the remaining 70% is spent in loading/unloading, positioning, measuring, exchanging cutters, idling, and so on (Ham, 1985).

The cost of small-lot-sized production is about 10-30 times larger than large-lot-sized production (He, 1987). Therefore, major efforts should be made to improve this situation for higher manufacturing productivity. The application of GT in manufacturing is to group similar parts into part families. Small-lot-sized production of parts in the same part families is collected into large-lot-sized group production, so the advantages of large-lot-sized production are acquired by small-lot-sized production. GT has been defined to be, in essence, as a broad philosophy that is aimed at the identification of part families, based on similarities in design and/or manufacturing features; and the systematic exploitation of these similarities in every phase of the manufacturing operation (Suresh and Kay, 1998). An overview of various elements of GT is provided in Figure 1.

It shows that through the identification and formation of part families, as well as cellular manufacturing, many synergistic benefits will be realized in the design stage, process planning stage, integration of design and process planning functions, production management stage, and in other stages downstream.

With GT to guide designing in the design stage instead of the traditional design approach, parts' design becomes more rational and more standard.

Figure 1. Elements of group technology



Parts are classified into design families according to similar criteria, which can be retrieved by designers at the design stage. The statistics shows that above 3/4 of designing new parts can refer to or directly excerpt from original ones (Xiang and Cai, 1987). Avoidance of "reinvention of the wheel" when designing new parts reduces the developmental lead times and costs. GT endows designing new parts with more similarities, which provides the foundations for GT implementing in manufacturing.

The preliminary applications of GT are in process planning, beginning with group working procedures. Parts waiting for machining are grouped into process families according to similar machining methods and installation methods. Process planning, which is suited to all parts in a process family, is carried out, hence setting-up times are reduced.

In production management, GT is a basic technology for computer-aided management systems. The computer is a very useful tool for collecting, analyzing, and processing large quantities of information in the process of managing and controlling production. Information is classified into groups, then normalized and standardized according to the similarity criteria. In this way, the efficiency of production management has been greatly improved.

PART FAMILY FORMATION: CODING AND CLASSIFICATION SYSTEMS

Coding is defined as a technique of allocating predetermined symbols to provide for meaningful information. In the coding and classification systems of GT, the code is the identifier that is used to identify the design and process features of parts. Coding is an effective approach to converting part information into a shorthand notation which is easy for computers to store, retrieve, and manipulate. Though GT may be practiced without coding systems, it is an essential and effective tool for successful implementation of GT philosophy, especially in the implementation of Computer Integrated Manufacturing (CIM).

Code Structures

Through coding, the descriptions for the attributes of complex parts are converted into symbol expressions, to achieve the goal of "symbols instead of shape". Full advantage can be taken of the computer's capabilities for information processing to statistically analyze part spectrums with coding, and to group into part families according to the requirements of structure-process similarity. A code may be numbers or letters, or a combination of numbers and letters, which are assigned to the parts for information processing.

Three basic structures of codes are used in classification and coding systems.

Monocodes (Hierarchical Codes)

A monocode or hierarchical code utilizes a tree structure and each node in the sequence is dependent on the selection made at the previous node. Therefore a monocode can be used to rapidly subdivide a population into small groups with this tree structure. However, it is much more difficult to interpret a given digit in this code because the left digit value of the given digit has to be considered. This may make monocoded systems less useful for computer analysis, although monocoded systems are ideally suited to the identification and retrieval of similar designs.

Polycodes

Unlike a monocode, each digit of a polycode has a meaning of its own and can be interpreted without consideration of any other digits. Polycodes are easy to construct and modify as required. Polycoded systems are generally suited for computer analysis. However, polycodes may become a longer code yet having only similar accuracy to monocodes because of the lack of logic relationship among their digits. Most of polycodes are of fixed length, but there are some of variable length.

Hybrids

Most classification and coding systems used in industry are hybrids of monocoded and polycoded systems. A common manner with hybrids is to divide the population into small groups using one or two monocode digits and then placing a polycode series in each branch. In this manner the benefits of both code structures can be combined in one code.

In general, through these three basic structures of codes, there are three basic attributes to be considered in grouping parts: (1) shape, (2) function, and (3) manufacturing operations and tooling.

Coding and Classification Systems

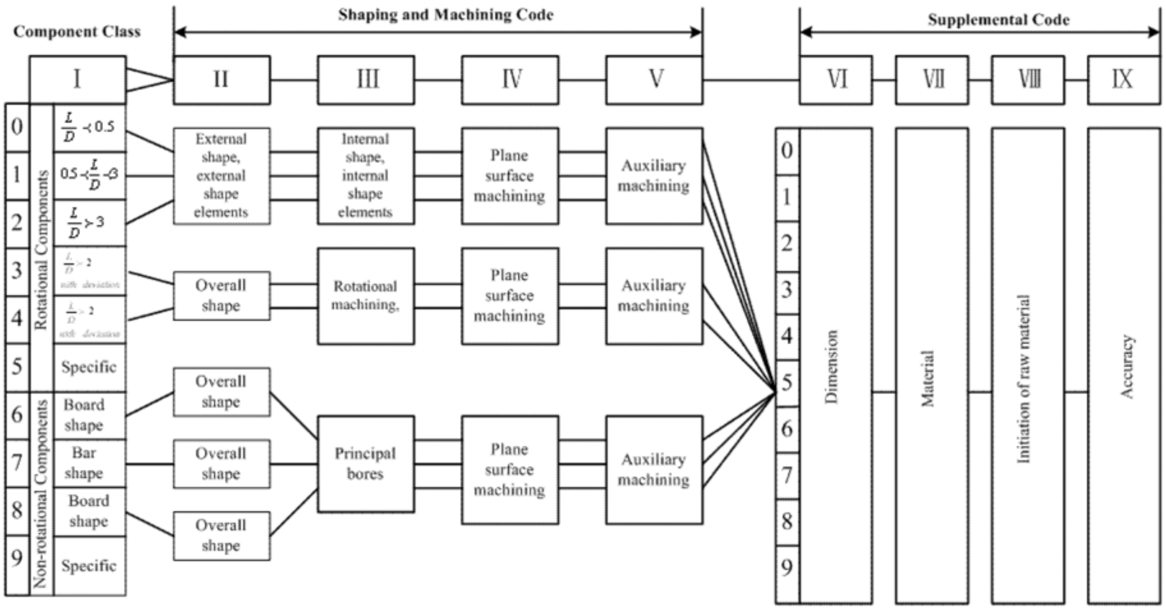
A well-designed and properly adapted classification and coding system is a key element for acquisitionof GT benefits, and the adaptation and implementation of a classification and coding system for GT applications is an important and complex task. Although many different kinds of classification and coding systems have been developed and used around the world, each company should source or develop a system suited to its needs and requirements. Three systems are presented in some detail in the following section to provide an insight into the types of systems available and to indicate the implementation of classification and coding systems.

The OPTIZ System

The OPTIZ system, developed by Professor H. Opitz in West Germany, is a decimal hybrid classification and coding system consisting of nine digits, of which five digits are basic and four digits are supplemental. The overall code structure is shown in Figure 2. The first digit is monocode used to divide the population of parts into major families. The second digit to further define the shape factor is quasi-polycode. The third to the fifth digits are polycode in their respective branches. The four digits in the supplemental code are used to classify size, material type, raw material initial form, and tolerance class.

Although the OPTIZ system emphasizes the structure features of parts, it implies somewhat the process features from the arrangement of the structure: from external shape, to internal shape, to plane surface machining, and to auxiliary machining. The OPTIZ system utilizes a digit to express accuracy, but the accuracy is a complicated problem: it includes dimensional accuracy, geometric accuracy, and positional accuracy, so the accuracy

Figure 2. The structure of OPTIZ code



cannot be completely depicted. From the view of its whole structure the OPTIZ system is quite simple, but from the view of its local structure it is quite complicated, because the digits in the code of the OPTIZ system are pertinent. As a whole, the OPTIZ system is simple, but not complete.

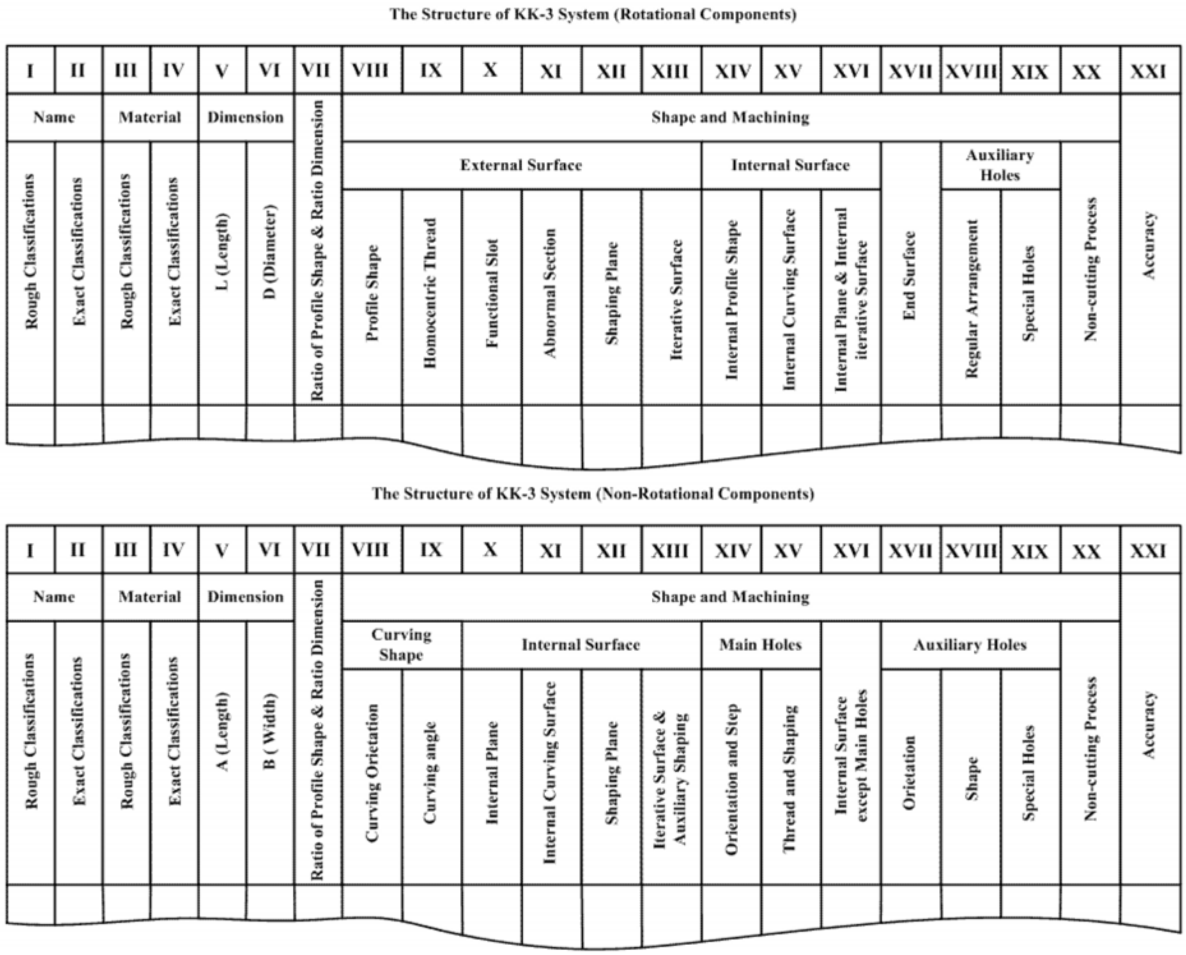
The KK-3 System

The KK-3 system was first presented in Japan in 1976 and is the third version of the KK system. Before the KK-3 system were the KK-1 system and the KK-2 system. The KK-1, presented in 1970, is a decimal hybrid system consisting of thirteen digits. The KK-2 presented in 1973 is a decimal hybrid system consisting of fifteen digits. The KK-3 system is also a decimal hybrid coding system consisting of twenty-one digits. Like the OPTIZ system, the KK-3 system firstly divides the population into two classes: rotational components and non-rotational components (Figure 3). Compared with the OPTIZ system, the KK-3 system is more complete, not only considering structure features but also process features. Of its code composed of twenty-one digits there is seven coterminous digitals at the front to be used to express design information, which is different from the OPTIZ system; the codes related to design information are dispersed, consequently the system is suited for design information retrieval. The KK-3 system has some shortcomings, such as the fact that the code is too long, and the probability for some digits to be used for component coding is very low.

The JLBM-1 System

The JLBM-1 system was presented in China in 1984. The JLBM-1 is a decimal hybrid classification and coding system consisting of fifteen digits. The JLBM-1 system is similar to the OPTIZ system in its structure, and incorporates some advantages of the of KK-3 system (Figure 4). In order to avoid the shortcomings of OPTIZ, the JLBM-1 system improves on the OPTIZ system, in that it extends the shape and machining code of the OPTIZ sys-

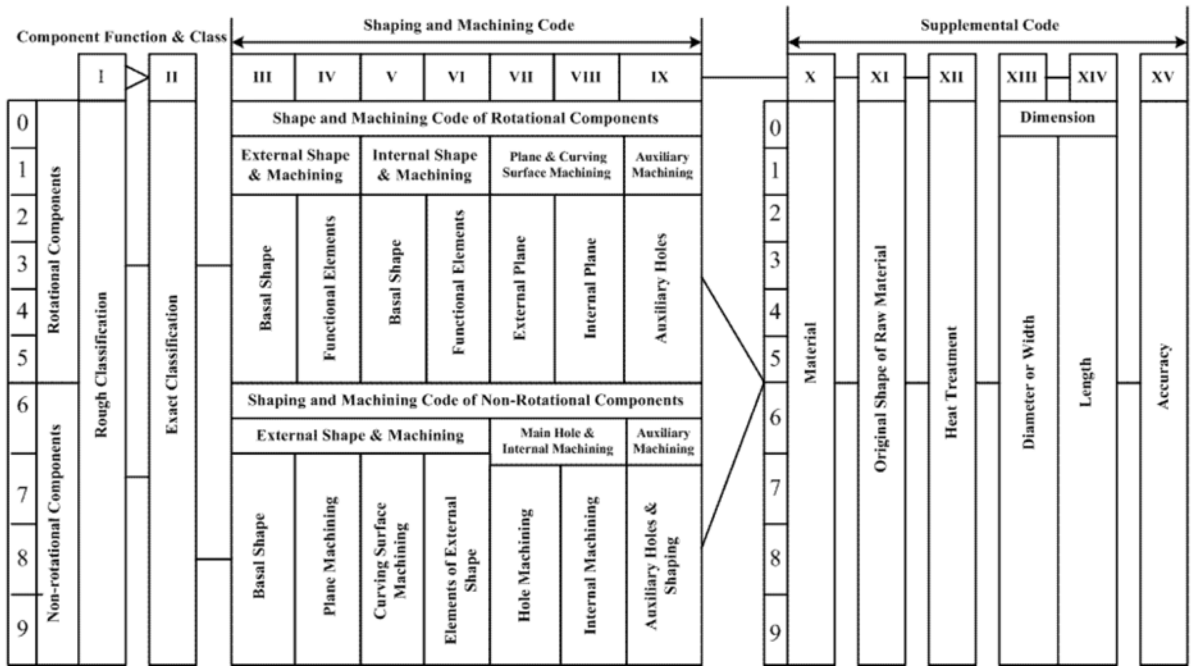
Figure 3. The structure of KK-3 system



tem; it uses the code of the component function and name instead of the code of the component class in OPTIZ; it separates the code of heat treatment from the material and heat treatment code of the OPTIZ system; it uses two digits to express the main dimension code instead of the single digit in the OPTIZ system. Although the JLBM-1 system has been presented by incorporating some advantages of both the OPTIZ system and the KK-3 system, there are some shortcomings in this system. Almost all information related to designing disposes at the end of its code, which is quite steady and usually does not vary. However the information related to shaping is not quite as steady and often varies in the middle of its code. Once the shape code has to expand, the position of the supplemental code will be changed. Like the shape code, if those codes which often need to expand or shorten are changed, the positions of the steady codes are also changed.

Although many classification and coding systems have been developed, no system has been found to be universally suited to all types of companies. Many companies, therefore, use their own in-house designed systems for their particular applications.

Figure 4. The structure of JLBM-1 system



Part Family Formation

Classification or part formation is a process of grouping similar parts or separating dissimilar parts based on predetermined attributes. For example, the parts may be classified on the basis of geometric shapes, dimensions, type of material, operation and so forth. The classification approaches are used to group similar parts into a part family based on their design features or processing features. Part families are those part collections with similar characteristics.

To group parts into families, all parts are assigned codes according to the coding system. These codes (representing parts) are then processed to form groups that are interpreted as part families. Each part p can be assigned a vector $\mathbf{X_p}$ of attribute values (Han & Ham, 1986).

$$\mathbf{X_p}=[x_{p1}, x_{p2}, \ldots, x_{pk}],\ i = 1, 2, \ldots, k$$

where k is the number of the digits of the coding and classification system. The similarity of between two parts q and r can be determined by a distance function which can be defined in many different ways. The most commonly applied distance metrics are the following.

Minkowski distance d_{qr}:

$$d_{qr} = \left[\sum_{i=1}^{k} | x_{qi} - x_{ri} |^{m} \right]^{1/m} \qquad (1)$$

where m is any real number larger than or equal to 1. Two special cases of the above metric which are widely use are the Euclidean distance metric (m=2) and Manhattan distance metric (m=1). The distance measures for all parts yields the symmetric matrix shown as a diagonal matrix (8.2). There are a number of cluster analysis methods to be used to group the parts. The objective of cluster analysis is to assign P parts to f part families while minimizing some measure of distance. Here are

three cluster analysis approaches to forming part families.

Multi-Objective Cluster Analysis Approach

This method was proposed by Han and Ham in 1986 for identifying flexible part families,

1 2 … *k*

$$[d_{qr}] = \begin{matrix} 1 \\ 2 \\ \vdots \\ k \end{matrix} \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1k} \\ d_{21} & \ddots & & d_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ d_{k1} & d_{k2} & \cdots & d_{kk} \end{bmatrix} \tag{2}$$

where 'flexible' is used in the sense that the user has the choice of input digit priority and similarity of digit set. The objective for this cluster analysis is to lexicographically minimize the distance between the digits. Minimize:

$$\left[\sum_{r=1}^{f}\sum_{q=1}^{P}(\gamma_{qr}d_{qrk})\right] \quad k \in Z \tag{3}$$

subject to:

$$d_{qrk}=0,\ \forall k \in z, p \text{ for all parts in part family } r \tag{4}$$

$$\sum_{r=1}^{f}\gamma_{qr} = 1, \quad for\ q = 1,2,\ldots,P \tag{5}$$

$$\gamma_{qr}=0,1 \tag{6}$$

Z is the classification code prioritized sequence and z is the set of digits of significant similarity.

Hierarchical Cluster Analysis Approaches

There are two kinds of hierarchical techniques: the agglomerative and divisive. They construct their hierarchy in opposite directions, possibly yielding quite different results (Kaufman & Rousseeuw, 1990).

Agglomerative methods start when all objects are apart (that is, at step 0 there are P clusters), so each part can be considered as forming a small cluster by itself. The method, which proceeds by a series of successive fusions of parts, is subsequently applied. At the step 1, the two "most similar" parts are merged to form a cluster of two parts while all other parts remain apart. Then in step two, clusters are merged, until the required part clusters are attained. The dissimilarity $d(R,Q)$ between the clusters R and Q is defined as the average of all dissimilarities dqr, where q is any part of cluster R and r is any part of cluster Q. Then more formally,

$$d(R,Q) = \frac{1}{|R||Q|}\sum_{\substack{q\in R \\ r\in Q}} d_{qr} \tag{7}$$

Divisive methods are hierarchical in nature. At the beginning all parts together form a cluster. Then at each step, a divisive method splits up a cluster into two smaller ones, until finally the required part clusters are attained.

P-Median Model

This is a mathematical programming approach to the cluster analysis problem. The objective of this model is to identify f part families optimally, such that the distance between parts in each family is minimized with respect to the median of the family. In the following notation P is the number of parts and f is the number of part families to be formed. Minimize:

$$\sum_{q=1}^{P}\sum_{r=1}^{P}\gamma_{qr}d_{qr} \tag{8}$$

subject to:

$$\sum_{r=1}^{P} x_{qr} = 1 \quad (9)$$

$$\sum_{q=1}^{P} x_{qq} = f \quad (10)$$

$$x_{qr} \leq x_{qq} \quad (11)$$

$$x_{qr} = 0 \text{ or } 1 \quad (12)$$

Equation (8) ensure that each part p is assigned to exactly one part family. The number of part families to be formed is specified by Equation (9). Apparently the number of part families to be formed is a parameter in this model. Equation (10) ensure qth part family to be formed only if the corresponding part is a median. And the last Equation (11) ensure integrality.

Part-Machine Formation

A manufacturing system based on the GT concept is known as the cellular manufacturing system (CMS). In a cellular manufacturing system, parts are classified into part families based on the similarities of their process requirements, and machines are grouped into machine cells based on the operation requirements of the part families, which is a procedure of part-family formation. The aforementioned part family formation emphasized the capture of part attributes, namely similar function, shape and so on, but gave no help in identifying the set of machines to process them. Production flow analysis (PFA) was proposed to find a complete division of all parts into families and also a complete division of all existing machines into associated groups by analyzing information in the process routes for parts (Burbidge, 1971,1989).

Production Flow Analysis

Production flow analysis is a technique which is used to simplify the material flow system and to find part families and machine groups for group plane layout (Burbidge, 1971). This approach is concerned principally with existing methods of manufacture but does not consider other features such as material type and form. Groups are formed through the analysis of operation route sequence data, for all the operations used in the plant. Production flow analysis usually has four stages: factory flow analysis, group analysis, line analysis, and tooling analysis. These four steps are progressively sub-techniques of production flow analysis, and each should be completed before the next can start.

Factory Flow Analysis

Factory flow analysis is concerned with the material flow among the processing units, incorporating these processing units into some major divisions, and searching the quite simple material flow system of these major divisions.

Group Analysis

Group analysis is concerned with the major division of facilities into groups and the major division of parts into associated families. Each part family can be completed inside only one facility group. The primary aim is to achieve the simplest possible material flow system inside each major division.

Line Analysis

Line analysis is concerned with the material flow between machines inside the facility groups, and with planning the best layout for machines. The aim is to find the sequence of machines which will give the nearest approximation to line flow.

Tooling Analysis

Tooling analysis is concerned with the division of process equipments into groups. The aim is to find the same process equipments and the optimal

Table 2. A production flow analysis chart

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	√	√		√	√		√	√	√		√	√		√	√		√	√	√	√	√	√	√		√
2	√	√	√		√	√	√		√		√		√	√		√				√	√			√	
3			√	√				√		√		√	√		√		√	√	√		√				
4	√	√	√	√		√	√	√		√	√	√	√	√		√	√	√		√	√	√		√	√
5	√	√	√	√		√			√			√	√		√			√		√			√	√	√

sequence of product inside facility groups. Table 2 is an example of a production flow analysis chart, which is a machine-part incidence matrix indicating the part numbers along the top and indicating the machine number along the left. There are 25 components in this example. Figure 5 is this production flow analysis chart after rationalization.

The classification of components with production flow analysis is quite complicated if the number of components is very large. In order to replace manual processing with the computer there are many approaches to production flow analysis. Two approaches, namely cluster analysis and artificial-neural-network-based analysis are described next.

Traditional Cluster Analysis

Cluster analysis is a kind of numerical classification which is concerned with the grouping of parts into homogenous clusters (families) based on design and/or processing features. The approach provides a way to study the similarities of the diverse populations of objects in a quantitative manner.

To illustrate the clustering concept, consider the machine-part incidence matrix (13). The machine-part incidence matrix $[a_{ij}]$ consists of 0, 1 entries, where an entry 1(0) indicates that machine i is used (not used) to process part j.

Part number 1 2 3 4 5 6

$$[a_{ij}] = \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \text{Machine number} \qquad (13)$$

Clustering algorithms allows the transformation of an initial incidence matrix which clusters of machines and parts are not visible into a more structured (possibly block diagonal) form.

Similarity Coefficient Methods

There are two coefficient methods to be discussed: the single linkage cluster analysis and the average linkage clustering approach. The single linkage cluster analysis (SLCA) is based on the similarity coefficient s_{ij} measure between two machines i and j and is computed as follows (McAuley, 1972):

$$s_{ij} = \frac{\sum_{k=1}^{n} \delta(a_{ik}, a_{jk})}{\sum_{k=1}^{n} \theta(a_{ik}, a_{jk})} \qquad (14)$$

where

$$\delta(a_{ik}, a_{jk}) = \begin{cases} 1 & if \quad a_{ik} = a_{jk} = 1 \\ 0 & otherwise \end{cases}, \quad \theta(a_{ik}, a_{jk}) = \begin{cases} 0 & if \quad ai_{k} = a_{jk} = 1 \\ 1 & otherwise \end{cases},$$

n = the number of parts.

To solve the grouping problem using SLCA approach, similarity coefficients for all possible

Figure 5. The production flow analysis chart after rationalization

	01	02	20	07	11	14	09	05	22	04	12	18	08	17	25	19	23	15	03	13	06	24	16	10	21
1	√	√	√	√	√	√	√	√	√																
2	√	√	√	√	√	√	√	√		Group 1															
4	√	√	√	√	√	√			√																
5	√	√	√				√																		
1										√	√	√	√	√	√	√	√	√							
3									Group 2	√	√	√	√	√		√		√							
4										√	√	√	√	√	√										
5										√	√	√			√		√	√							
2														2 Family					√	√	√	√	√		√
3																		Group 3	√	√				√	√
4																			√	√	√	√	√	√	√
5																			√	√	√	√			
																							3 Family		

pairs of machine are computed. Machine groups are generated based on the threshold value of the similarity coefficient.

To illustrate the SLCA, consider matrix (8.12). The similarity coefficients s_{ij} are computed and the similarity coefficient matrix is obtained from the similarity coefficients.

$$s_{12} = s_{14} = s_{23} = s_{34} = \frac{0}{6} = 0$$
$$s_{13} = \frac{3}{3} = 1$$
$$s_{24} = \frac{2}{4} = 0.5$$

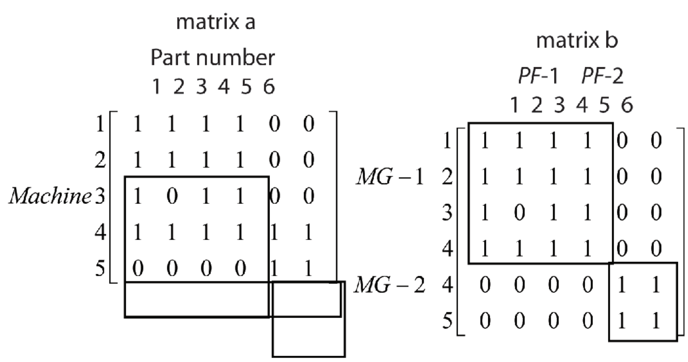
Assuming the threshold value of the similarity coefficient s_{ij}=50%, the following machine groups and corresponding part families are obtained (see Figure 6).

One of the disadvantages of SLCA is that it fails to recognize the chaining problem that results from the duplication of bottleneck machines. A bottleneck machine is a machine that does not allow decomposition of a machine-part incidence matrix into a submatrix. For example, machine 4 in Figure 7a does not permit decomposition of that matrix into two machine groups and two part families. A way to decompose Figure 7a into two mutually separable submatrices is to add an additional copy of machine 4. The Figure 7a is transformed into Figure 7b. Two machine groups, *MG*-1={1, 2, 3, 4} and *MG*-2={4, 5}, and two corresponding part families, *PF*-1={1, 2, 3, 4} and *PF*-2={5, 6}.

Figure 6.

$$\begin{array}{ccc} & & \begin{array}{cccccc} \multicolumn{3}{c}{PF\text{-}1} & \multicolumn{3}{c}{PF\text{-}2} \\ 2 & 4 & 5 & 3 & 6 & 1 \end{array} \\ MG-1 & \begin{array}{c} 1 \\ 3 \end{array} & \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ \hline \end{array}\right. \\ MG-2 & \begin{array}{c} 2 \\ 4 \end{array} & \left.\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right] \end{array}$$

Figure 7. Matrix a and Matrix b



The average linkage clustering (ALC) approach is applied to overcome the chaining problem (Seifoddini and Wolfe, 1986). The similarity coefficient between any two clusters is defined as an average of the similarity coefficient between all members of the two clusters. To solve the group problem using ALC the machine-part incidence matrix is represented using a binary machine code. The total number of intercellular movements (ICM) between two machine groups *MC-i* and *MC-j* is computed as follows:

$$ICM_{ij} = \sum_{k=1}^{n} \delta(v_{ik}, v_{jk}) \qquad (15)$$

where

$$m_{ik} = \begin{cases} 1 & if \quad \sum_{i \in MC-i} a_{ik} \neq 0 \\ 0 & otherwise \end{cases}$$

$$\delta = \begin{cases} 1 & if \quad m_{ik} = m_{jk} = 1 \\ 0 & otherwise \end{cases}$$

Note that $\delta(v_{ik}, v_{jk})$=1 indicates that part *k* requires processing in both machine tools *MC-i* and *MC-j*. This problem can be solved either by removing part *k* from the two machine groups or by adding an identical machine to each machine group (like matrix (8.15)). Consider the example illustrated in matrix (16), the intercellular movement is computed as follows:

$$\begin{matrix} MG-1 \\ \\ MG-2 \\ \end{matrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \qquad (16)$$

$[m_{1k}]$=[1, 1, 0, 0, 1]

$[m_{2k}]$=[0, 0, 1, 1, 1]

$$ICM_{12} = \sum_{k=1}^{5} \delta(m_{1k}, m_{2k}) = 0 + 0 + 0 + 0 + 1 = 1$$

The only part that requires processing in both machine groups *MG*-1 and *MG*_2 is part 5.

Rank Order Clustering

The rank order cluster (ROC) is a method of clustering based on sorting rows and columns of the machine-part incidence matrix (King, 1980).

ROC Algorithm

- **Step 1.** For each row and column of the machine-part incidence matrix, assign binary weight and calculate a decimal equivalent

(a weight). A weight for each row *i* and column *j* are calculate as follows:

Row *i*: $\sum_{k=1}^{n} a_{ik} 2^{n-k}$

Column *j*: $\sum_{k=1}^{m} a_{kj} 2^{n-k}$

- **Step 2.** Sort rows and column of the binary matrix in decreasing order of the corresponding decimal weights;
- **Step 3.** Repeat the preceding steps until the position of each element in each row and column does not change.

Another interesting sort-based algorithm is the direct clustering algorithm (DCA) (Chan and Milner, 1982).

DCA Algorithm

- **Step 1.** Determine the total number of ones in each row and column in the machine-part incidence matrix;
- **Step 2.** Sort each row in increasing order corresponding to the total number of ones;
- **Step 3.** Sort each column in decreasing order corresponding to the total number of ones;
- **Step 4.** Repeat the preceding steps until the position of each element in each row and column does not change;

Cluster Identification Algorithm

The cluster identification (CI) algorithm allows one to check the existence of mutually separable clusters in a machine-part incidence matrix, provided that they exist (Kusiak & Chow, 1987).

CI Algorithm

- **Step 1.** Set iteration number k=1;
- **Step 2.** Select any row *i* of incidence matrix $A^{(k)}$ and draw horizontal line h_i through it, where $A^{(k)}$ represents the matrix *A* at iteration *k*;
- **Step 3.** For each entry of one crossed by the horizontal line h_i draw a vertical line v_j;
- **Step 4.** For each entry of one crossed once by a vertical line v_j draw a horizontal line h_k.
- **Step 5.** Repeat steps 2 and 3 until there are no more crossed-once entries of 1 in $A^{(k)}$. all crossed-twice entries of one in $A^{(k)}$ from machine group MG-k and part family PF-k;
- **Step 6.** Transform the incidence matrix $A^{(k)}$ into $A^{(k+1)}$ by removing rows and columns corresponding to all the horizontal and vertical lines drawn in steps 1 through 4;
- **Step 7.** If matrix $A^{(k+1)}$=0, stop; otherwise set k=k+1 and go to step 1.

Consider machine-part incidence matrix (8.17) cannot be separated into two disjoint submatrices because of part 5, which is to be machined in two machine groups, *MG*-1 and *MG*-2. To deal with the overlapping parts, the extended CI algorithm was developed.

Extended CI Algorithm

- **Step 1.** Set iteration number k=1;
- **Step 2.** Select rows of matrix $A^{(k)}$ that are potential candidates for inclusion in machine group MG-k based on the user's expertise. Draw horizontal line h_i through each raw of matrix $A^{(k)}$ corresponding to these selected machines;
- **Step 3.** For each column in $A^{(k)}$ corresponding to the entry of one crossed by any of the horizontal line draw a vertical line v_j;
- **Step 4.** For each row in $A^{(k)}$ corresponding to the entry of one crossed by the vertical line v_j, drawn in step 2, draw a horizontal line h_i. A temporary machine group MG'-*k* is formed that is based on the machines

corresponding to all the horizontal lines drawn in step 1 and step 3. If some of machines cannot be included in the temporary machine group MG'-k based on the production manager's expertise, the corresponding horizontal lines in matrix $A^{(k)}$ are erased.

Delete from matrix $A^{(k)}$ parts (columns) that are to be manufactured on at least one of the machines already included in *MG-k*. Put these parts on the list of parts to be manufactured in a functional machining facility;

Draw a vertical line v_j through each crossed-once entry of one in $A^{(k)}$ which does not involve any other machines than those included in *MG-k*;

- **Step 5.** For all the crossed-twice entries of one in $A^{(k)}$, form a machine group *MG-k* and part family *PF-k*;
- **Step 6.** Transform the incidence matrix $A^{(k)}$ into $A^{(k+1)}$ by removing all the rows and columns corresponding to the rows and columns included in *MG-k* and *PF-k*, respectively;
- **Step 7.** If matrix $A^{(k+1)}=0$, stop; otherwise set $k=k+1$ and go to step 1.

If a user is not satisfied with the machine groups and part families generated. The entire computational process can be repeated. Initiating the extended CI algorithm with another machine (or machines) in step 2 may result in a different configuration of groups of machines and parts.

Artificial Intelligence (AI) based Cluster Analysis

Recognizing that traditional methods of grouping parts into families have their limitations and the need for incorporating some form of intelligence in the grouping process, researchers have proposed AI-based techniques for solving the GT problem.

Syntactic Pattern Recognition

Wu et al. (1986) first explored the use of a non-traditional technique for GT. They applied principles of syntactic pattern recognition from the formal language theory to design of manufacturing cells given the parts' machine sequences. Syntactic recognition has been commonly used in the field of AI for parsing strings in natural language processing. Machine sequences are treated as strings which are used to form part families. The method derives a dendrogram from which multiple clusterings with different number of cells can be derived. It is possible to generate alternative groupings of some parts from the dendrogram allowing the decision maker to balance load among cells to a certain extent. Classification of a new/modified part into an existing machine cell can be done by matching the part's machine sequence with that of cell's pattern.

Expert System

In the paper of Kusiak and Hergua (1987),they discussed the applications of expert systems to GT. To perform this task, Kusiak (1988) proposed a method based on expert systems, the most successful and develop branch of AI. Kusiak starts with the machine-part incidence matrix modified to include the processing time of part j on machine i instead of the binary (0,1) data. The procedure to form part families and machine cells uses a knowledge-base and a clustering algorithm closely interacting with each other. Each iteration determines machine cells and selects part families based on a heuristic which takes into account variables such as production costs, material carrier, etc. The assignment is checked for satisfaction of three or four meta-constraints such as availability of processing time in a cell and maximum number of machine cells. if any of the constraints is violated, appropriate rules from the knowledge-base are fired to correct the situation.

The problems of AI based techniques are that they are not capable of generalization if a new input does not fall into the existing domain specific knowledge; and it is difficult to acquire new knowledge.

Computational Intelligence (CI) Based Cluster Analysis

Artificial Neural Network Approach

Artificial neural networks have emerged as a useful addition to the set of tools and techniques for the application of group technology. The application of artificial neural networks for the part-machine group problem has produced very effective results.

ART-1 Approach

As the foregoing the ART-1 approach, it can be used to classify a set of binary vectors into groups, each group containing similar vectors based on a prespecified degree of similarity. The ART-1 network architecture consists of two layers of neurons, the input layer or comparison layer with p neurons, and the output layer or recognition layer with q neurons, where $\mathbf{W_j}=[w_{1,j}, w_{2,j}, \ldots, w_{p,j}]$ is a weight vector and t_j is an exemplar pattern (see Figure 8). The number of neurons p at the input layer is usually the dimension of the binary input vectors that are to be classified. The comparison layer usually receives three inputs: an external binary input vector $\mathbf{X}=[x_1, x_2, \ldots, x_p]$ characterizing a part or a machine; a feedback vector $\mathbf{T}=[t_1, t_2, \ldots, t_q]$ from the output layer; and a gain vector $\mathbf{G}=[g_1, g_2, \ldots, g_p]$ The output of the neurons at the comparison layer is 1 if any two of the three inputs to it is 1 (called the two-third rule for activation). The steps involved in a serial implementation are summarized below.

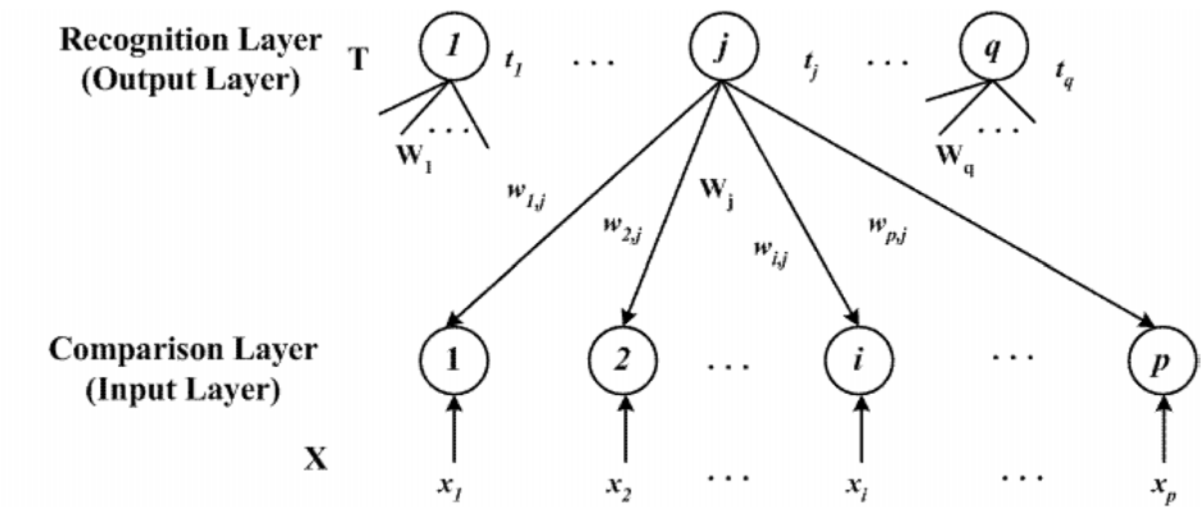
- **Step 1.** Initialization
  Set $g_i = 1$ and $t_j(1)=0$ for input nodes i=1, 2, .., p and output nodes j=1, 2, …, q. Select a valued for vigilance threshold ρ: $0 \leq \rho \leq 1$.
- **Step 2.** Read new input vector **X**, consisting of 0/1 elements x_i.
  Compute matching scores of output t_j of every output node j.
  $t_j = \sum_{i=1}^{p} w_{ij}(t) i_i$ for i=1, 2, …, p
- **Step 3.** Select best matching exemplar: $t_{k*} = \sum_j \max\{t_j\}$; and $t_j = 0$ ($j \neq k*$),suppress outputs of other neurons.
- **Step 4.** Vigilance test (i. e. degree of similarity with best-matching exemplar)

$$similarity = \frac{number\ of\ matching\ 1s\ in\ input\ and\ exemplar}{number\ of\ 1s\ in\ X} \quad (17)$$

  If *similarity* > ρ, go to step 7; else go to step 6.
- **Step 5.** Disable best-matching exemplar temporarily
- **Step 6.** Update best-matching exemplar (fast update mode)
  g_i=0, $w_{i,k*} = \dfrac{Lc_i}{L - 1 + \sum_{k=1}^{p} c_k}$
  where c_i is the ith component of the comparison layer output vector, $k*$ is the number of the winning recognition layer neuron, and L is a constant that usually set to 2 or 3.
- **Step 7.** Repeat: go to step 2, after enabling any nodes disabled in step 6.

ART-1 in its form has many drawbacks (Malavé & Ramachandran, 1994). The obtained classification is very sensitive to the sequence in which the vectors are applied during training. Another shortcoming is that the above approach does not offer a way to deal with bottleneck machines and their duplication. Finally, the determination of

Figure 8. ART-1 neural network architecture



the optimal vigilance parameter is not straightforward.

An Improved ART-1 Approach

The improved ART-1 approach presented by Dagli and Huggahalli (1992) incorporates a few changes in the basic approach. The improved ART-1 approach involves the following steps:

a. Find the optimal vigilance values for a particular matrix size.
   One technique involves finding an optimal vigilance value to ensure that the number of machine and part groups would be similar, if not equal.
b. Obtain the final part-machine matrix using the optimal vigilance values
c. Separate part and machine groups. Identify each block..
d. Find the blocks with highest degree clustering, defined by the following formula.

$$\frac{number\ of\ 1s\ within\ a\ block}{\dim ension\ of\ the\ block}$$

The improved ART-1 approach addresses the issue of obtaining optimal vigilances, and detecting and automatically duplicating a bottleneck machine. It is also an approach that integrates the procedures for part family and machine cell formation, detection and duplicating bottleneck machines, and cost performance analysis. This approach leads to an integrated approach for optimal part-machine group formation.

Genetic Algorithm

As the aforementioned GA is a well-known structured random search method. An important virtue of GA is that it will tend to converge on solutions that are globally optimal, or nearly so, even in a large and complicated search space under certain conditions on the problem domain. The purpose of part-machine formation is to find the best combination of machine cells and part families under some physical and logical constraints so that specific objectives are achieved. Building up a mathematical model is therefore the first action in solving problems with GA. The mathematical model of part-machine formation can be conveniently formulated using the following notations (Chan, Mak, Luong, & Ming, 1998):

- i = the machine number;
- j = the part number;

- k = the cell number;
- M = the total number of machines;
- P = the total number of parts;
- L = the maximal number of manufacturing cell;
- x_{ik} = binary variable indicating whether the machine I is assigned to the cell k;
- y_{ik} = binary variable indicating whether the part j is assigned to the cell k;
- C_v, C_e = constant variable used as coefficient;
- N_v = total number of voids inside the diagonal blocks;
- N_e = total number of exceptional elements outside the diagonal blocks;
- Z = the objective function value.

The objective function is to minimize

$$Z=C_vN_v+C_eN_e \tag{18}$$

where

$$N_v = \sum_{k=1}^{L}[(\sum_{i=1}^{M} xik)(\sum_{j=1}^{P} yjk)] \tag{19}$$

$$N_e = \frac{\sum_{i=1}^{M}\sum_{j=1}^{P} a_{ij}(\sum_{k=1}^{L}\left|x_{ik} - y_{jk}\right|)}{2} \tag{20}$$

And subject to the following constraints:

$$C_v+C_e=1 \tag{21}$$

$$\sum_{k=1}^{L} x_{ik} = 1 \text{, for } i\in\{1, 2, 3,\ldots, M\} \tag{22}$$

$$\sum_{k=1}^{L} y_{jk} = 1 \text{, for } j\in\{1, 2, 3,\ldots, P\} \tag{23}$$

where

$$x_{ik} = \begin{cases} 1 & machine\ i\ is\ assigned\ to\ cell\ k \\ 0 & otherwise \end{cases} \tag{34}$$

$$y_{jk} = \begin{cases} 1 & machine\ i\ is\ assigned\ to\ cell\ k \\ 0 & otherwise \end{cases} \tag{25}$$

In order to solve the machine-component grouping problem by using the technique of genetic algorithms, it should be specially designed in accordance with the nature of the problem, including individual representation, fitness evaluation, parent selection, reproduction and mutation. The steps of GA are as follows.

- **Step 1.** Individual representation (Chromosome)
  Here an integer string is employed in order to improve the efficiency of the algorithm (see Figure 9).
- **Step 2.** Initialization
  Select the initial input parameters (population size, number of generations, probabilities of crossover, mutation, and number of cells) and generate randomly an initial diversified population.
- **Step 3.** Selection and recombination
  The Equation (18) is used as the function to evaluation the fitness of each individual. By recording the best (minimum) value of objective function, which is noted as Z_b in each generation, the fitness value of each individual p in generation g, with the objective function value of Z_p, can be calculated as follows:

  $$\Gamma^{(p)} = \frac{Z_b}{Z_p} \tag{26}$$

  The fitness value, given in Equation (25), indicates that the individual with a lower

Figure 9. Individual representation

Individual (Chromosome)							
Machines				Parts			
x_1	x_2	...	x_M	y_1	y_2	...	y_P

value of its objective function has a higher fitness value.
The parents are selected in accordance with the principle that the individual with a higher fitness value will be selected more often

- **Step 4.** Crossover/recombination
  The two point crossover operation is employed to generate new individuals. One cross point is selected randomly within the genes representing machines, and another cross point within the genes representing parts.
  The mutation operation is performed by selecting randomly an individual from the population.
- **Step 5.** Termination
  The generation will continue to evolve until a specified stopping criterion has been satisfied.

Fuzzy Approach

The foundation of fuzzy approach is fuzzy set theory (FST). According to the foregoing FST, the basic theory of FST is classes with unsharp boundaries. FST is suitable for dealing with impreciseness that may exist in the parameters of any system. Impreciseness practically exits in grouping parts and machines.

Fuzzy Classification

Conventional methods, as mentioned before, implicitly assume that disjoint groups exist in the data set and therefore a part (machine) can belong to only one group. The grouping result can thus be expressed as a binary matrix: $G=[u_{ij}]_{c\times n}$, such as equation (8.15), where c is the number of groups; n is the number of parts (machines). In fuzzy classification, suppose that n parts are to be grouped into g families. However, space u_{ij} is not restricted to binary value 0 and 1, u_{ij} is a value between 0 and 1, and $\sum_{i=1}^{m} u_{ij} = 1$ that is, a part may belong to several part families at the same time with different weights. The steps in implementing the fuzzy classification method is as follows (Xu, 1989).

- **Step 1.** Part-family formation.
  - Select a sample part set, determine the features which are relevant to the machining process under consideration, and define a membership function for each selected feature.
  - Calculate the similarity matrix $\mathbf{S}_{n\times n}$.
    Suppose that the sample consists of n parts and every part has p features. The ith part can be represented by a vector $X_i=(x_{i1}, x_{i2}, \ldots, x_{ip})$. Let us use the fuzzy number $\mu_k(x_{ik})$ ($i=1, 2, \ldots, n$; $k=1, 2, \ldots, p$) to denote the membership of the ith part to the kth feature. The similarity $s(X_i, X_j)$ of two parts can be calculated directly by the following formulas (Li, Ding, & Lei, 1988).

$$s(X_i, X_j) = \frac{\sum_{k=1}^{p} \min(\mu_k(x_{ik}), \mu(x_{jk}))}{\frac{1}{2}\sum_{k=1}^{p} (\mu_k(x_{ik}) + \mu_k(x_{jk}))} \tag{27}$$

$$s(X_i, X_j) = \frac{\sum_{k=1}^{p} \mu_k(x_{ik}) \cdot \mu(x_{jk})}{(\sum_{k=1}^{p} (\mu_k(x_{ik})^2 \cdot \mu_k(x_{jk})^2))^{1/2}} \tag{28}$$

$$s(X_i, X_j) = 1 - \frac{\sum_{k=1}^{p} | (\mu_k(x_{ik}) - \mu(x_{jk}) |}{\sum_{k=1}^{p} \max(\mu_k(x_{ik}), \mu_k(x_{jk}))} \quad (29)$$

$$s(X_i, X_j) = 1 - \frac{\sum_{k=1}^{p} | (\mu_k(x_{ik}) - \mu(x_{jk}) |}{\sum_{k=1}^{p} (\mu_k(x_{ik}) + \mu_k(x_{jk}))} \quad (30)$$

The coefficients of similarity have the following properties:
$0 \leq s(X_i, X_j) \leq 1$ for $i \neq j$
$s(X_i, X_j) = s(X_j, X_i)$
$s(X_i, X_i) = 1$

- Specify a proper value of λ or c. A λ value may be considered as a similarity threshold for the parts or a c value is the number of part-family in order to be in the same part family.
- Fuzzy classification.
  Here $V_i = (v_{i1}, v_{i2}, \ldots, v_{ip})$ is defined as the mean value of the parts in the ith part family and V_i is considered to be the conceptual reference part of the ith part family.

$$v_{ik} = \frac{\sum_{j=1}^{n} u_{ij} \cdot \mu_k(x_{jk})}{\sum_{j=1}^{n} u_{ij}}, \; i=1, 2, \ldots, c; \; k=1, 2, \ldots, p \quad (31)$$

- **Step 2.** Assign new parts into part families.
  - Calculate the feature values for each new part and convert the values into fuzzy numbers.
  - Fuzzy pattern recognition
    Since c reference patterns, $V_1, V_2, \ldots, V_c$ are already obtained, the new part X is classified into ith reference pattern if the following condition is satisfied:

$$(X, V_i) = \bigvee_{1 \leq k \leq c} (X, V_k) \succ t \quad (32)$$

where (X, V_i) denotes the closeness of two fuzzy sets and t is the given threshold value; (X, V_i) is calculated by the following formulas.

$$(X, V_i) = \frac{1}{2}((\bigvee_{x \in X_i} (\mu_X(x) \wedge \mu_{V_i}(x))) + (1 - \bigwedge_{x \in X} (\mu_X(x) \vee \mu_{V_i}(x)))) \quad (33)$$

If this closeness value is smaller than the threshold value t, then put it in a dummy part family for later processing.

This fuzzy classification can generate quite realistic results through the selection of features based on the characteristics of the machining process. Either a similarity parameter λ or the desired number of part families c can be used to control the part-family formation process and modifying a membership function dynamically can assure more realistic.

Fuzzy ART Network (FART)

Here the ideas of FST are imported into the ART network. Fuzzy ART involves several changes to ART1, including: (1) non-binary (analogue) input vectors can also be processed; (2) there is now a single weight-vector connection (w_{ij}); (3) in addition to ρ, two other parameters need to be specified: a choice parameter (α) and a learning rate parameter (β); (4) a general updating rule is utilized, which provides another effective means for countering category proliferation. The steps involves in FART are as follows (Suresh, Slomp, & Kaparthi, 1999):

- **Step 1.** Initialization
  Select values for: Choice parameter: $\alpha > 0$; Learning rate: $\beta \in [0,1]$ and vigilance parameter: $\rho \in [0,1]$.
- **Step 2.** Read new input vector ***I*** which is the normalized vector of the input vector **X** and

consists of binary or analogue elements.

- **Step 3.** Compute choice function (T_j) for every output node
  $$T_j = \frac{\left\| I \wedge W_j \right\|}{\alpha + \left\| W_j \right\|}$$
  for nodes j=1, 2, …, q(34)
  where $\wedge$ is the fuzzy AND operator.
- **Step 4.** Select best-matching exemplar: $T_{k*} = \max_j \{T_j\}$.
- **Step 5.** Resonance test
  If $Similarity = \frac{\left\| I \wedge W_{k*} \right\|}{\left\| I \right\|} \geq \rho$ go to learning step 7.
  Else go to the next.
- **Step 6.** Mismatch reset: Set T_{k*}=-1 and go to step 4.
- **Step 7.** Update best-match exemplar
  $$W_{k*}^{new} = \beta(I \wedge W_{k*}^{old}) + (1-\beta)W_{k*}^{old} \qquad (35)$$
- **Step 8.** Repeat: go to step 2.

This FART neural network involves three parameters to be specified, including the choice parameter α, the learning rate β, and the vigilance parameter ρ. The main problem of FART is the category proliferation problem because of the general learning rule in which the exemplar vectors are updated in the light of newly presented inputs. In order to solve this problem, Özdemir *et al.* (2007) modified the FART algorithm, in which the fuzzy MIN operator ($\wedge$) of the learning rule was replaced by fuzzy MAX ($\vee$) operator and also presented a two-stage clustering approach with modified FART.

GROUP TECHNOLOGY IN INTELLIGENT MANUFACTURING

In order to achieve the higher productivity and efficiency in multi-product, small-lot-sized production and mass customization of products, it is absolutely essential to incorporate the concept of GT into every manufacturing activity including operations and management.

Group Technology in CAD

GT has been recognized as a means to reduce design redundancy since its inception. GT can be applied through design standardization and rationalization techniques to reduce the proliferation and bring about increased standardization and rationalization. With the advent and development of computer and network technologies, they provide computer aided design (CAD)/computer aided manufacturing (CAM) are provided with significantly enabled technologies. It has been recognized that GT is an essential element of the foundation for the successful development and implementation of CAD/CAM through application of the part-family concept based on geometrical and processing similarities between parts. Under the guidance of GT, product design can not only take full advantage of available design information, but can also endow the designed product with more similarities. This establishes a foundation for the implementation of GT in the sequence of production activities.

Group Technology in Process Planning

Process planning is generated by process planners who use their knowledge and experience based on the design, the available facilities and manpower of actual manufacturing procedures for producing products. Sometimes, new equipment and tools are required or the design is subcontracted to other companies. Process planning has been defined by the Society of Manufacturing Engineers (Tulkoff, 1987) as "the systematic determination of the methods by which a product is to be manufactured economically and competitively". Process plans are generally represented by operation sheets or routers in which detailed operation procedures, parameters and facilities are specified.

Computer-aided process planning (CAPP) has been a very popular research topic for the last three decades. Many researchers have been working towards commercial software for carrying out process planning tasks or largely assisting the human process planner. There are two major approaches to computer-aided process planning: variant and generative. For relatively well-established companies, the variant approach may be useful since such companies have relatively stable products. The generative approach is advantageous for one-of-a-kind production, small volume, continuous new product development, or for contracting companies which commonly deal with new products [Gu, 1995].

Variant Process Planning

Most of the existing processes planning systems currently used in industry are based on the variant approach. The variant process planning procedure consists of two phases, the preparation phase and the production phase.

In the preparation phase, two tasks are performed: generation of the GT code for each component and generation of the process route for each component or for complex parts. Both items are stored in the database. Generation of the GT code is based on a previously created classification and coding system. Generally, all of them divide parts according to basic shape, dimensions, detailed shape features, material, and tolerances. The next task in the preparation phase is the generation of process plans. This is usually performed by copying the data from process routes created by experts into the computer database.

In the production phase three tasks are performed: generation of the GT code for the new part, retrieval of the process route for a similar part in the computer database from the computer database, and adjustment of this similar part for the new part. The modified plan thus becomes a process plan for the new part which is stored back into the database. This procedure may be repeated for each new part in the production phase of the process planning system.

The benefits of these systems were that the manufacturing and process planning knowledge of experienced human planners was captured in the form of process plans. There was no need to repeat the planning procedure as it was performed in a uniform way. The main problem with this approach was that rapid and significant changes in the design of new products and parts would not allow for the existence of a similar process plan, and therefore planning for such parts had to start from the beginning, from the design.

Generative Process Planning

Generative process planning is an approach that received research attention in the early 1980s by the infusion of artificial intelligence methods into manufacturing planning (Sormaz, 1998). It starts with a part design (which should include the complete design specification, i.e., geometry, dimensions, tolerances, etc.), then generates the process plan based on existing manufacturing knowledge, and finally provides a full description for these processes and generates the optimal sequence of processes for manufacturing. A truly generative process planning system is a turnkey system that creates process plans based on built-in decision logic without human intervention (Chang et al. 1991).

The knowledge base appears a very important part of generative CAPP systems. The capture of manufacturing process information in a systematic way may also be seen as an application of GT in a broad sense. For example, the representation of machining processes as objects with attributes and methods for process selection, tool path generation, and so on in an algorithmic way allows its uniform application for all parts, and therefore is indeed GT in its broadest sense.

Grouping Scheduling

The scheduling problem, focusing on the determination of the sequence or order in which the part families assigned to a cell should be processed in order to either minimize or maximize some measure of performance, is generally known as group scheduling (Lognedran, 1998). Such group scheduling procedures which exploit similarities in setups existing among the individual parts within a given part family can reduce overall machine setup time. To implement group scheduling procedures the parts within the part family are first partitioned into subfamilies based on setup similarities.

The group scheduling approach typically encompasses two stages. The first stage involves sequencing jobs within each subfamily. These decisions are made by utilizing traditional job shop scheduling rules. The second stage consists of determining which subfamily queue to select, and when to switch to another subfamily queue at each machine. These decisions are addressed by queue selection heuristics.

According to the existing group scheduling literatures, group scheduling rules can be classified as either exhaustive or non-exhaustive. Exhaustive heuristics are such that once a subfamily is chosen, all the jobs within that subfamily, including new arrivals, are processed before another subfamily can be considered. Only when there are no jobs left in that subfamily can jobs in other subfamilies be processed. Conversely, non-exhaustive heuristics allow switching to another subfamily even when the current subfamily queue is not empty. While exhaustive rules tend to minimize setups and are simpler to implement, some non-exhaustive rules may enhance due date performance (Mahmoodi & Dooley, 1991).

SUMMARY

It is now four decades since these early pioneers of GT first described their ideas. GT ideas have come a long way since then. As discussed in this chapter, GT plays a very important role in the transformation from mass production to multi-product, small-lot-sized production. The most appealing benefits of GT are the lower material handling costs, the reduced work-in-process, and the shorter setup times. In this chapter, the foundation and principles of GT have been introduced. The approaches to part family formation and manufacturing cellular layout have been stressed. Overall, there are two different approaches to part family formation and manufacturing cellular layout; one is based on the particular features of the parts, and the other is based on the production flow of the process. GT applications in CAD, process planning, and grouping scheduling have been discussed. With changes to the manufacturing environment and the development of computer and network technology, GT should be further developed to adapt to these changes. Some new approaches (such as artificial neural network, fuzzy logic, and genetic algorithm) are used in GT, which provides part-machine formation with effective approaches. As an organizational approach which justifies small and medium batch sized production systems, GT is exploited in various phases of manufacturing and demonstrates its advantages. Further improvements in GT will offer the potential for enhanced impact on future.

REFERENCES

Burbidge, J. L. (1971). Production Flow Analysis. *Production Engineering*, *4*(5), 139–152.

Burbidge, J. L. (1989). *Production Flow Analysis for Planning Group Technology*. Oxford, UK: Oxford Science Publications, Clarendon Press.

Chan, F. T. S., Mak, K. L., Luong, L. H. S., & Ming, X. G. (1998). Machine-Component Grouping Using Genetic Algorithms. *Robotics and Computer-integrated Manufacturing*, *14*, 339–346. doi:10.1016/S0736-5845(98)00024-6

Chan, H. M., & Milner, D. A. (1982). Direct Clustering Algorithm for Group Formation in Cellular manufacturing. *Journal of Manufacturing Systems*, *1*(1), 65–74. doi:10.1016/S0278-6125(82)80068-X

Chang, T. C., Wysk, R. A., & Wang, H. P. (1991). *Computer-Aided Manufacturing*. Upper Saddle River, NJ: Prentice Hall.

Chu, C. H., & Tsai, C. C. (1998). A Heuristic Genetic Algorithm for Manufacturing Cell Formation . In Kamrani, A. K., & Logendran, R. (Eds.), *Group Technology and Cellular Manufacturing: Methodologies and Applications*. New York: Gordon and Breach Science Publishers.

Dagli, C. H., & Huggahalli, R. (1992). A Neural Network Approach to Group Technology . In Wong, J., & Yoshiyasu, T. (Eds.), *Neural Networks in Design and Manufacturing*. Singapore: World Scientific Pub. Co.

Gallagher, C. C., & Knight, W. A. (1986). *Group Technology Production Methods in Manufacture*. New York: Halsted Press.

Gu, P., & Norrie, D. H. (1995). *Intelligent Manufacturing Planning*. London: Chapman & Hall.

Ham, I., Hitomi, K., & Yoshida, T. (1985). *Group technology Applications to Production Management*. Amsterdam: Kluwer-Nijhoff Publishing.

Han, C., & Ham, I. (1986). Multiobjective Cluster Analysis for Part Family Formations. *Journal of Manufacturing Systems*, *5*(4), 223–230. doi:10.1016/0278-6125(86)90053-1

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

King, J. R. (1980). Machine-Component Group Formation in Production Analysis: An Approach Using a Rank Order Clustering Algorithm. *International Journal of Production Research*, *18*(2), 213–232. doi:10.1080/00207548008919662

Kusiak, A. (1988). EXGT-S: A Knowledge Based System for Group Technology. *International Journal of Production Research*, *26*(5), 887–904. doi:10.1080/00207548808947908

Kusiak, A. (1990). *Intelligent Manufacturing Systems*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Kusiak, A., & Chow, W. S. (1987). An Algorithm for Cluster Identification. *IEEE Transactions on Systems, Man, and Cybernetics*, *17*(4), 696–699. doi:10.1109/TSMC.1987.289363

Kusiak, A., & Hergau, S. S. (1987). Group Technology. *European Journal of Operational Research*, *29*(3), 229–251. doi:10.1016/0377-2217(87)90238-4

Logendran, R. (1998). Group Scheduling for Cellular Manufacturing Systems . In Kamrani, A. K., & Logendran, R. (Eds.), *Group Technology and Cellular Manufacturing: Methodologies and Applications*. New York: Gordon and Breach Science Publishers.

Mahmoodi, R., & Dooley, K. J. (1991). A Comparison of Exhaustive and Non-Exhaustive Grouping Scheduling Heuristics in a Manufacturing Cell. *International Journal of Production Research*, *29*, 1923–1939. doi:10.1080/00207549108948058

Malavé, C. O., & Ramachandram, S. (1994). Machine-Part Family Formation . In Dagli, C. H. (Ed.), *Artificial Neural Networks for Intelligent Manufacturing*. Boca Raton, FL: Chapman & Hall.

McAuley, J. (1972). Machine Grouping for Efficient production. *Production Engineering*, *2*, 53–57.

Seifoddini, H., & Wolfe, P. M. (1986). Application of the Similarity Coefficient Method in Group Technology. *IEEE Transactions*, *18*(3), 271–277. doi:10.1080/07408178608974704

Shunk, D. (1985). Group Technology Provides Organized Approaches to realizing benefits of CIMS. *Industrial Engineering (American Institute of Industrial Engineers)*, *17*, 416–423.

Sormax, D. N. (1998). GT and CAPP: Towards An Integration of Variant and Generative Approaches . In Suresh, N. C., & Kay, J. M. (Eds.), *Group Technology and Cellular Manufacturing: State-of-the-Art Synthesis of Research and Practice*. Boston: Kluwer Academic Publishers.

Suresh, N. C., & Kay, J. M. (1998). *Group Technology and Cellular Manufacturing: State-of-the-Art Synthesis of Research and Practice*. Boston: Kluwer Academic Publishers.

Suresh, N. C., Slomp, J., & Kaparthi, S. (1999). Sequence-Dependent Clustering of Parts and Machines: A Fuzzy ART Neural Network Approach. *International Journal of Production Research*, *37*(12), 2793–2816. doi:10.1080/002075499190527

Tulkoff, J. (1987). Process Planning: A Historical Review and Future Prospects. In *Proceedings of the 19th CIPR International Seminar on Manufacturing System Computer-Aided Process Planning*, Pennsylvania State University, (pp. 207-210).

Xu, H., & Wang, H. (1989). Part Family Formation for GT Application Based on Fuzzy Mathematics. *International Journal of Production Research*, *27*(9), 1637–1651. doi:10.1080/00207548908942644

Xu, X. H., & Cai, J. G. (1987). *Group Technology*. Peking, China: Mechanical Industry Press.

Chapter 9
Intelligent Control Theory and Technologies

In the previous chapters, the theoretical and technological foundations of MI have been investigated, they include, knowledge-based system, multi-agent system, data mining and knowledge discovery, computing intelligence, process and system modeling, sensor integration and data fusion, and group technologies. As another very important branch of MI, intelligent control theory and technology will be discussed in this chapter. The chapter should be viewed as a resource for those in the early stages of *considering* the development and implementation of intelligent controllers for industrial applications. It is impossible to provide the full details of a field as large and diverse as intelligent control in a single chapter. Hence, the focus is placed on the fundamental ideas that have been found most useful in industry.

DOI: 10.4018/978-1-60566-864-2.ch009

INTRODUCTION

Intelligent control is a discipline where the control methods developed attempt to emulate important characteristics of human intelligence. These characteristics include adaptation and learning, planning under large uncertainty and coping with large amounts of data (Antsaklis 1997). Today, the area of intelligent control tends to encompass everything that is not characterized as conventional control. Intelligent control is interdisciplinary as it combines and extends theories and methods from areas such as control, artificial intelligence, computer science and information technology and operations research. It uses theories from mathematics and seeks inspiration and ideas from biological systems. Intelligent control methodologies are being applied to robotics and automation, communications, manufacturing, traffic control, to mention but a few application areas.

Adaptive control, fuzzy control, neural network control, knowledge-based control, expert system control, hybrid system control as well as learning control are all areas where related work is taking place (Antsaklis & Passino 1993, Antsaklis et al. 1996). The areas of computer science and in particular artificial intelligence provide knowledge representation ideas, methodologies and tools such as semantic networks, frames, reasoning techniques and computer languages such as Lisp and Prolog. Concepts and algorithms developed in the areas of adaptive control and machine learning help make intelligent controllers adaptive and smart. Advances in sensors, actuators, computation technology and communication networks provide the necessary for implementing intelligent controllers.

Why is intelligent control needed? The fact is that there are problems of control today that cannot be formulated and studied using conventional differential/difference equation mathematical model or "conventional control" methodologies which were developed in the past decades to control dynamical systems (Antsaklis 1994). To address these complex problems in a systematic way, a number of methods have been developed in recent years that are known as "intelligent control" methodologies. Intelligent control uses conventional control methods to solve "lower level" control problems and conventional control is included in the area of intelligent control (Alur & Dill 1994). Intelligent control attempts to build upon and enhance the conventional control methodologies to solve new challenging control problems.

The chapter will be organized as follows. The chapter will begin with an overview of the foundations for intelligent control, which includes the difference between conventional control and intelligent control, the concepts of intelligence and intelligent control, the definition of intelligent system, and the notions of control and intelligent system. The models for intelligent controllers will be presented and the relative intelligent control technologies such as fuzzy control, neural network control and learning control will then be described. In addition, intelligent control systems include hierarchical control system, expert control system, neural control system, integrated intelligent control system, and network based intelligent control system are investigated. The chapter will also explain in broad terms as to apply the methods to challenging problems. Finally, a conclusion will be drawn from the chapter.

FOUNDATIONS OF INTELLIGENT CONTROL

The term "intelligent control" has come to mean, particularly to those outside the control area, some form of control using fuzzy and/or neural network methodologies (White & Sofge 1992). Intelligent control, however does not restrict itself only to those methodologies. In fact, according to some definitions of intelligent control not all neural/fuzzy controllers would be considered intelligent. The fact is that there are problems of control today that cannot be formulated and studied using conventional differential/difference equation mathematical framework and "conventional control" methodologies; these methodologies were developed in the past decades to control dynamical systems (Albus 1991). To address these problems in a systematic way, a number of methods have been developed in recent years that are known as "intelligent control" methodologies. There are significant differences between conventional and intelligent control. It is worth remembering at this point that intelligent control uses conventional control methods to solve "lower level" control problems, and conventional control is included in the area of intelligent control. In summary, intelligent control attempts to build upon and enhance the conventional control methodologies to solve new challenging control problems.

Conventional and Intelligent Control

The word control in "intelligent control" has different meanings, it is more general than the word control in "conventional control". First, the processes of interest are more general and may be described, for example by either discrete event system models or differential/difference equation models or both. This has led to the development of theories for hybrid control systems, which study the control of continuous-state dynamic processes by discrete-state controllers. In addition to the more general processes considered in intelligent control, the control objectives can also be more general. For example, "replace part A in satellite" can be a general task for the controller of a space robot arm; this is then decomposed into a number of subtasks, several of which may include for instance "follow a particular trajectory", which may be a problem that can be solved by conventional control methodologies. To attain such control goals for complex systems over a period of time, the controller has to cope with significant uncertainty that fixed feedback robust controllers or adaptive controllers cannot deal with. Since the goals are to be attained under large uncertainty, fault diagnosis and control reconfiguration, adaptation and learning are important considerations in intelligent controllers. It is also clear that task planning is an important area in the design of intelligent controllers. The control problem in intelligent control is an enhanced version of the problem in conventional control. It is much more ambitious and general. It is not surprising that these increased control demands require methods that are not typically used in conventional control. The area of intelligent control is in fact interdisciplinary, and it attempts to combine and extend theories and methods from areas such as control, computer science and operations research to attain demanding control goals in complex systems.

Note that the theories and methodologies from the areas of operations research and computer science cannot, in general, be used directly to solve control problems, as they were developed to address different needs; they must first be enhanced and new methodologies need to be developed in combination with conventional control methodologies, before controllers for very complex dynamical systems can be designed in systematic ways. Conventional control concepts such as stability may have to be redefined when, for example, the process to be controlled is described by discrete event system models; and this issue is being addressed in the literature. Concepts such as reachability and deadlock developed in operations research and computer science are useful in intelligent control, when studying planning systems. Rigorous mathematical frameworks, based on predicate calculus are being used to study such questions. However, in order to address control issues, these mathematical frameworks may not be convenient and they must be enhanced or new ones must be developed to appropriately address these problems. This is not surprising as the techniques from computer science and operations research are primarily analysis tools developed for non real-time systems, while in control, synthesis techniques to design real-time feedback control laws for dynamic systems are mainly of interest. In view of this discussion, it should be clear that intelligent control research, which is mainly driven by applications that have a very important and challenging theoretical component. Significant theoretical strides must be made to address the open questions. The problems are nontrivial, but the pay-off is very high indeed.

As it was mentioned above, the word control in intelligent control has a more general meaning than in conventional control; in fact it is closer to the way the term control is used in every day language. Because intelligent control addresses more general control problems that also include the problems addressed by conventional control, it is rather difficult to come up with meaningful bench mark examples. Intelligent control can address control problems that cannot be formulated in the language of conventional control.

To illustrate, in a rolling steel mill, for example, while conventional controllers may include the speed (rpm) regulators of the steel rollers, in the intelligent control framework one may include in addition, fault diagnosis and alarm systems; and perhaps the problem of deciding on the set points of the regulators, that are based on the sequence of orders processed, selected based on economic decisions, maintenance schedules, availability of machines etc.. All these factors have to be considered as they play a role in controlling the whole production process which is really the overall goal.

Another difference between intelligent and conventional control is in the separation between controller and the system to be controlled. In conventional control the system to be controlled, called the plant, typically is separate and distinct from the controller. The controller is designed by a control designer, while the plant is in general given and cannot be changed; note that recent attempts to coordinate system design and control have been reported in areas such as space structures and chemical processes, as many times certain design changes lead to systems that are much easier to control. In intelligent control problems, which are most often complex and challenging, there may not be a clear separation of the plant and the controller; the control laws may be imbedded and be part of the system to be controlled. This opens new opportunities and challenges as it may be possible to affect the design of processes in a more systematic way.

Areas relevant to intelligent control, in addition to conventional control, include hybrid systems, planning and knowledge based systems, machine learning, search algorithms, fault diagnosis and control reconfiguration, predicate logic, automata, Petri nets, neural nets and fuzzy logic. In addition, in order to control complex systems, one has to deal effectively with the computational complexity issue; this has been in the periphery of the interests of the researchers in conventional control, but it is clear that computational complexity is a main issue whenever one attempts to control complex systems.

Intelligence and Intelligent Control

It is appropriate at this point to briefly comment on the meaning of the word intelligent in "intelligent control". Note that the precise definition of "intelligence" has been eluding mankind for thousands of years. More recently, this issue has been addressed by disciplines such as psychology, philosophy, biology and of course by artificial intelligence (AI); note that AI is defined to be the study of mental faculties through the use of computational models. No consensus has emerged as yet of what constitutes intelligence. The controversy surrounding the widely used IQ tests, also points to the fact that we are well away from having understood these issues. This chapter introduces and discusses several characterizations of intelligent systems that appear to be useful when attempting to address complex control problems.

Intelligent controllers can be seen as machines which emulate human mental faculties such as adaptation and learning, planning under large uncertainty, and coping with large amounts of data in order to effectively control complex processes; this is the justification for the use of the term intelligent in intelligent control, since these mental faculties are considered to be important attributes of human intelligence. An alternative term is "autonomous (intelligent) control" which emphasizes the fact that an intelligent controller typically aims to attain higher degrees of autonomy in accomplishing and even setting control goals, rather than stressing the (intelligent) methodology that achieves those goals. We should keep in mind that "intelligent control" is only a name that appears to be useful today. In the same way the "modern control" of the 60's has now become "conventional control", as it has become part of the mainstream, what is called intelligent control today may be called just "control" in the not so distant future.

What is more important than the terminology used are the concepts and the methodology, and whether or not the control area intelligent control will be able to meet the ever increasing control needs of our technological society.

Definition of Intelligent Control Systems

Intelligent systems can be characterized in a number of ways and along a number of dimensions. There are certain attributes of intelligent systems that are of particular interest in the control of systems. We begin with a general characterization of intelligent systems: An intelligent system has the ability to act appropriately in an uncertain environment, where an appropriate action is required to increase the probability of success, and success is the achievement of behavioral goals that support the system's ultimate goal. In order for a man-made intelligent system to act appropriately, it may emulate functions of living creatures and ultimately human mental faculties.

An intelligent system can be characterized along a number of dimensions. There are degrees or levels of intelligence that can be measured using the various dimensions of intelligence. At a minimum, intelligence requires the ability to sense the environment, make decisions and control related action. Higher levels of intelligence may include the ability to recognize objects and events, represent knowledge in a model, and plan for the future. In advanced forms, intelligence provides the capacity to perceive and understand, choose wisely, and act successfully under a large variety of circumstances so as to survive and prosper in a complex and often hostile environment. Intelligence can be observed to grow and evolve, both through growth in computational power and through accumulation of knowledge of how to sense, decide and act in a complex and changing world.

The above characterization of an intelligent system is rather general. According to this, a great number of systems can be considered intelligent. In fact, according to this definition even a thermostat may be considered to be an intelligent system, although of low level of intelligence. It is common however to call a system intelligent when in fact it has a rather high level of intelligence. There exist a number of alternatives but related definitions of intelligent systems which emphasize systems with high degrees of intelligence. For example, the following definition emphasizes the fact that the system in question processes information, and it focuses on man-made systems and intelligent machines: Machine intelligence is the process of analyzing, organizing and converting data into knowledge; where (machine) knowledge is defined to be the structured information acquired and applied to remove ignorance or uncertainty about a specific task pertaining to the intelligent machine. This definition relates to the principle of increasing precision with decreasing intelligence of Saridis.

Next, an intelligent system can be characterized by its ability to dynamically assign subgoals and control actions in an autonomous fashion: Many adaptive or learning control systems can be thought of as designing a control law to meet well-defined control objectives. This activity represents the system's attempt to organize or order its "knowledge" of its own dynamical behavior, so to meet a control objective. The organization of knowledge can be seen as one important attribute of intelligence. If this organization is done autonomously by the system, then intelligence becomes a property of the system, rather than of the system's designer. This implies that systems which autonomously (self)-organize controllers with respect to an internally realized organizational principle are intelligent control systems.

A procedural characterization of intelligent systems is given next: Intelligence is a property of the system which emerges when the procedures of focusing attention, combinatorial search, and generalization are applied to the input information in order to produce the output. One can easily

deduce that once a string of the above procedures is defined, the other levels of resolution of the structure of intelligence are growing as a result of the recursion. Having only one level structure leads to a rudimentary intelligence that is implicit in the thermostat, or to a variable-structure sliding mode controller.

Control and Intelligent Systems

The concepts of intelligence and control are closely related and the term "Intelligent control" has a unique and distinguishable meaning. An intelligent system must have defined goals. Control is then required to achieve these goals. Consequently, any intelligent system will be a control system. Conversely, intelligence is necessary to provide desirable functioning of systems under changing conditions, and it is necessary to achieve a high degree of autonomous behavior in a control system. Since control is an essential part of any intelligent system, the term "intelligent control systems" is sometimes used in engineering literature instead of "intelligent systems" or "intelligent machines". The term "intelligent control system" simply stresses the control aspect of the intelligent system.

A control system consists of data structures or objects (the plant models and the control goals) and processing units or methods (the control laws). An intelligent control system is designed so that it can autonomously achieve a goal, while its components, control goals, plant models and control laws are not well defined, either because they were not known at the design time or changed unexpectedly.

MODELS FOR INTELLIGENT CONTROLLERS

In highly autonomous control systems, the plant is sometimes so complex that it is either impossible or inappropriate to be described by conventional mathematical system models consisting of only differential or difference equations. Even though it might be possible to accurately describe some systems with highly complex nonlinear differential equations, such descriptions may be inappropriate if it makes subsequent analysis too difficult or too computationally complex to be useful. The complexity of the plant model needed in design depends on both the complexity of the physical system and on how demanding the design specifications are. There is a tradeoff between model complexity and our ability to perform analysis on the system via the model. Frequently, if the control performance specifications are not too demanding, a more abstract, higher level model can be utilized, which will make subsequent analysis simpler. This model intentionally ignores some of the system characteristics, specifically those that need not be considered in attempting to meet the particular performance specifications. For example, a simple temperature controller could ignore almost all dynamics of the house or the office and consider only a temperature threshold model of the system to switch a furnace off or on.

Discrete Event System Models

Discrete Event System (DES) models that use finite automata or Petri nets, queuing network models, Markov chains, etc. are useful for modeling the higher level decision making processes in an intelligent autonomous controller. The choice of whether to use such models will depend on what properties of the autonomous system are to be studied. More specifically, DES models are appropriate for general expert control systems, planning systems, abstract learning control and often the higher "management and coordination levels" in the hierarchical architecture for intelligent autonomous systems. DES analysis and controller synthesis techniques have been developed with remarkable success (Ramadge & Wonham 1989). Other important topics for intelligent control include approaches to controllability,

reachability, stability, and performance analysis. Applications of DES theoretic techniques have been reported for the modeling and analysis of AI planning systems and the stability analysis of expert control systems (Passino & Antsaklis 1993, Passino & Lunardhi 1996). Discrete event systems are important in their own right and they have been studied using many approaches. They are also very useful in connection with hybrid systems. Recently, an efficient methodology for supervisory controller design for DES was developed using Petri nets (Moody 1997, Moody & Antsaklis 1998). The approach uses the concept of place invariants of the net to design control supervisors that enforce linear constraints on the marking and firing vectors of the net. Potential applications of the approach in intelligent control include real-time control reconfiguration and planning different control tasks, for example in manufacturing and hybrid systems.

In general, when considering the application of DES theoretic techniques in intelligent control systems, it is important to study their computational aspects. Problems like reachability, liveness, and deadlock detection arise in many intelligent control applications. Studying the computational issues of DES approaches can be very important in automated verification, controller synthesis, on-line reconfiguration, and task planning among others. Several models have been proposed in literature to describe the dynamics of DES. An important observation is that higher expressiveness of the model typically results in algorithms of higher complexity. Petri nets are a trade off between expressiveness and complexity and are suitable for describing concurrent processes that appear frequently in intelligent systems. Petri nets are studied at length in this paper with respect to their computational properties.

Hybrid System Models

Hybrid systems are dynamical systems the behavior of interest of which is determined by interacting continuous and discrete dynamics (Antsaklis & Nerode 1998). These systems typically contain variables or signals that take values from a continuous set (e.g. the set of real numbers) and also variables that take values from a discrete, typically finite set (e.g. the set of symbols {a, b, c}). These continuous or discrete-valued variables or signals depend on independent variables such as time, which may also be continuous or discrete; some of the variables may also be discrete event driven in an asynchronous manner.

There are several reasons for using hybrid models to represent the dynamic behavior of interest. Reducing complexity was and still is an important reason for dealing with hybrid systems. This is accomplished by incorporating models of dynamic processes having different levels of abstraction (Alur et al. 1995, Petridis & Kaburlasos 1998).

A thermostat is a very simple example, but adequate enough to address the task of modeling complex heat flow dynamics. Another example, in order to avoid dealing directly with a set of nonlinear equations one may choose to work with sets of simpler equations (e.g. linear), and switch among these simpler models. The advent of digital machines has made hybrid systems very common. Whenever a digital device interacts with the continuous world, the behavior involves hybrid phenomena that need to be analyzed and understood.

Hybrid control systems typically arise from computer aided control of continuous industrial processes, manufacturing and communication networks for example. They also arise from the hierarchical organization of complex control systems. Hierarchical organization helps manage complexity and higher levels in the hierarchy require less detailed models (discrete abstractions) of the functioning of the lower levels (continuous dynamics), necessitating the interaction of discrete and continuous components. The study of hybrid control systems is essential in designing sequential supervisory controllers for continuous

systems, and it is essential in designing intelligent control systems with a high degree of autonomy. Hybrid system analysis and controller synthesis techniques can be used for design and verification of intelligent control systems.

Hybrid control systems appear in the intelligent autonomous control system framework whenever one considers the execution level together with control functions performed in the higher coordination and management levels (Branicky 1995). Examples include expert systems supervising and tuning conventional controller parameters, planning systems setting the points of local control regulators, and sequential controllers deciding which one of a number of conventional controllers is to be used to control a system, to mention but a few.

The analysis, design, simulation, and verification of hybrid systems require the development of computationally efficient algorithms and approaches. Several models have been proposed in the literature for the development of analysis and controller synthesis techniques. Timed automata and hybrid automata have been used by several researchers for modeling, verification and controller synthesis of hybrid systems. Although the initial results concerning the complexity of approaches based on timed and hybrid automata were negative, recent efforts have proposed systematic techniques that are applicable to a large class of problems. Because of the importance of hybrid automata based methods, we outline basic computational issues of hybrid automata based approaches later in the chapter.

Recently, a class of timed Petri nets named programmable timed Petri nets (Lemmon et al. 1998) has been used to model hybrid control systems. The main characteristic of the proposed modeling formalism is the introduction of a clock structure which consists of generalized local timers that evolve according to continuous-time vector dynamical equations. They can be seen as an extension of the approach (Alur et al. 1995) that provide a simple, but powerful way to annotate the Petri net graph with generalized timing constraints expressed by propositional logic formulae. It is expected that the more powerful expressiveness of Petri nets will result in analysis and controller synthesis approaches of higher complexity than those based on hybrid automata. However, there are complex systems which include concurrency and/or conflict, buffer sizes that can be modeled more compactly using Petri nets than finite automata. There are also control specifications, for example mutual exclusion constraints, that can be studied more efficiently in a Petri net framework. Moreover, there is the need to investigate the applicability of recent results in Petri nets in a hybrid framework. Stability and supervisory control design of hybrid systems modeled by programmable timed Petri nets have been studied in (Koutsoukos et al. 1998).

INTELLIGENT CONTROL TECHNOLOGIES

In this section we provide a brief overview of the main areas of intelligent control. The objective is not to provide a comprehensive in-depth review, but present the basic concepts of the approaches. The basic concepts and principles of intelligent control are proposed for complex control tasks. The main difference between conventional and intelligent control is that a conventional control method depends on the model of a system while intelligent control can address the control issues of non-model systems. The most promising development branches of intelligent control so far, are fuzzy control, neural network control and learning control.

Fuzzy Control

Fuzzy logic control theory was firstly introduced by Professor Zadeh in 1965. Its main principle lies on the idea of mimicking the fuzzy features of human thinking for the effective control of

uncertain systems through fuzzy logic reasoning (Mak & Wong 2000). The conventional control method is to control the target objects based on their mathematical models while fuzzy control seeks to control objects from the viewpoint of human intelligence. The core design idea of fuzzy control is the definition of fuzzy control rule and membership function. There are two important parts of fuzzy control, which are the development of fuzzy control principle and the realization of fuzzy control algorithm.

The membership functions and fuzzy control rules of a classic fuzzy logic controller are normally summarized based on prior experience of the system to be controlled. Though there is no amended function to the rule during the control process, and it does not have the ability of self-learning, fuzzy control is still widely used in broad fields such as stove control, chemical industry process control, water treatment, and household appliances (Celano et al. 2000). Meanwhile, a variety of improved or compound fuzzy controllers have also emerged, such as the fuzzy PID regulator, fuzzy expert controller, fuzzy adaptive controller, and fuzzy neural network controller. Currently, research efforts are placed on the modeling of fuzzy system, the stability and the robustness analysis of fuzzy controller, which will certainly boost the development of fuzzy control theory and technology and move fuzzy control into a new era.

Neural Network Control

Artificial neural networks are circuits, computer algorithms, or mathematical representations loosely inspired by the massively connected set of neurons that form biological neural networks. Artificial neural networks are an alternative computing technology that have proven useful in a variety of pattern recognition, signal processing, estimation, and control problems (Agogino et al. 2000).

Rosenblatt presented a special type of neural network in 1958, known as the Perception. It was intended to be a simple model for a life-form system to apperceive information from the outside. This model is mainly used for classification.

In 1982, U.S. physicist Hopfield proposed a kind of feedback network and defined an energy function as well. The energy function can be employed to define the status of neurons and the connection weight between neurons while the proposed network can be used to solving optimal calculation for associative memory. The network was later known as Hopfield network whose most typical application is to successfully resolve the optimal traveling path for tourism industry.

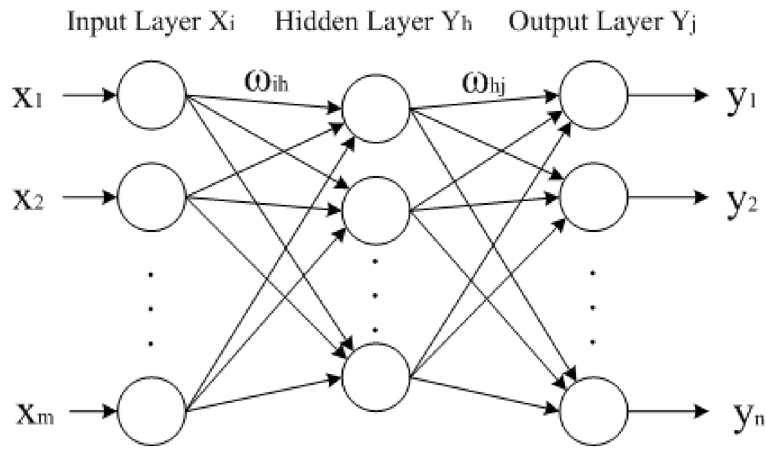
In 1986, Rumelhart and Mcclelland developed a Back Propagation (BP) algorithm for Multi-layer feed-forward neural network, which is called BP algorithm or BP network, to successfully solve the problem depicted briefly in Chapter 5.

As shown in Figure 1, the BP network consists of three layers which are the input layer, hidden layer, and output layer. The neurons between each layer are connected with different weight, and the network can be trained to discover the mapping relationships between input and output of the training samples. The achievement of Hornik indicated that BP network can be used to approximate any nonlinear function with arbitrary accuracy when increase the number of hidden layers and its neurons. The most typical BP network is the three-tier-forward-network, the structure of which is shown in Figure 1.

The Objective Function of BP Algorithm

The training mode of BP network belongs to supervisory training and the basic optimization method of BP network is the gradient descent method. The basic idea of BP algorithm is to provide the network with training samples including the input values and expected output values. By regulating the connection weights between layers and the threshold of neurons according to the deviation of the actual output values and the expected output ones, the goal is to make the actual outputs

Figure 1. The structure of BP network



approximate the expected outputs or at least fall into the range of permissible error.

The basic BP algorithm is based on the minimum quadratic form performance index function E which is defined as follows.

$$E = \sum_{k=1}^{P} E_k \tag{1}$$

where, P is the number of training samples and E_k is the local error function of k[th] training sample. E_k can be defined as

$$E_k = \sum_{j=1}^{n} \Phi(e_{j,k}) = \frac{1}{2}\sum_{j=1}^{n} e_{j,k}^2 = \frac{1}{2}\sum_{j=1}^{n} (y_{j,k} - \hat{y}_{j,k})^2 \tag{2}$$

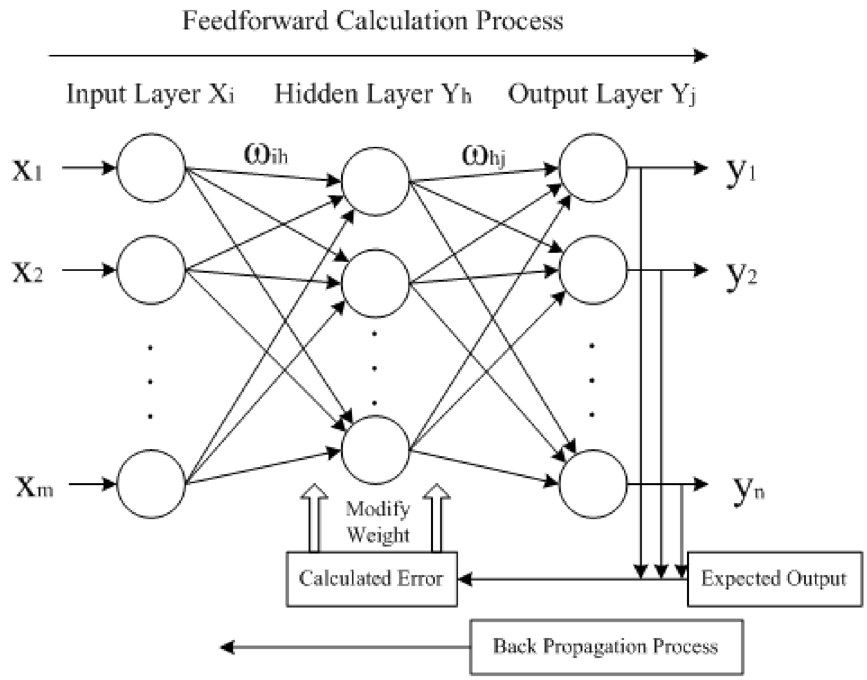
where, n is the node number of the output layer, $y_{j,k}$ is the expected output of k[th] samples derived from neuron j, $\hat{y}_{j,k}$ is the actual output of k[th] samples derived from neuron j, and $e_{j,k}$ is the error of k[th] samples derived from neuron j.

The Learning Process of BP Algorithm

As indicated in Figure 2, the learning process of BP network includes two stages: 1) the first stage is the feed-forward calculation process, which calculates the output of each node one by one until all the nodes of network are completed; 2) the second stage involves the error back propagation process, which calculates the connection weight for each node and make the corresponding modifications.

There are four operation processes during the two stages described above. The first one is called "mode forward propagation" process, which transmits the input signal from the input layer to the hidden layer and finally to the output layer of the BP network. The second process is known as "error back propagation" process that transmits the error signal (calculated value between the expected output and the actual output) from the output layer to the hidden layer and then to the input layer, and modifies the connection weights of each layer. The third one is the network "memory training" process that alternately perform the "mode forward propagation" process and the "error back propagation" process. The last one is "learning convergence" process, which reduces the global error of the network so that it falls within a defined range.

Figure 2. Learning process of BP network



Improved BP Algorithm

The BP algorithm is the most popular learning algorithm for multi-layer feed-forward neural network due to its stable theoretical basis, rigorous derivation process, symmetrical and graceful calculation formula, clear physical concept, and strong generality. Nevertheless, many disadvantages have also been discovered when using the BP algorithm to solve the practical problems, such as the network can easily fall into a local minimum value, the convergence of the learning process is slow, the number selection of both the hidden layer and its node lies on the human experience and lacks theoretical guidance, the BP network is a kind of feed-forward network without feedback connections, which affects the information exchange speed and efficiency.

In order to address the disadvantages mentioned above, an improved BP algorithm has been proposed by researchers (Zeng et al. 1998, Lei et al. 2006, Chen et al. 2008). The improved algorithm adds a inertia term (impulse term) to the regulation formula of the connection weight, to smooth the weight variation and prevent repeated oscillation without convergence. In the improved BP algorithm, when entering the kth samples, the regulation value of the connection weight between neuron *h* in the hidden layer and neuron *j* in the output layer is defined by Equation (3).

$$\Delta w_{hj,k} = \eta \cdot \varepsilon_{j,k} \cdot \hat{y}_{h,k} + \alpha \Delta w_{hj,k-1} \quad (3)$$

The regulation value of the connection weight between neuron *i* in the input layer and neuron *h* in the hidden layer is defined by Equation (4).

$$\Delta w_{ih,k} = \eta \cdot \varepsilon_{h,k} \cdot x_{i,k} + \alpha \Delta w_{ih,k-1} \quad (4)$$

In Equations (3) and (4), α is the inertia coefficient and $0 \le \alpha \le 1$, $\varepsilon_{j,k}$ and $\varepsilon_{h,k}$ are the inertia terms which can enhance the convergence rate and restrain the parasitic oscillation.

The neural network model is employed to simulate the activities of human brain's neuron, which includes the processing, handling, depositing and searching of information. It is characterized by the following features: 1) neural network stores information in a distributed way so that the information can be recovered even though the local network has been destroyed; 2) neural network can process and reason information in a parallel manner which means that each neuron in the network can independently operate and process the information received and transmit processing result to a destination; and 3) the information processing of neural network is characterized by self-organization and self-learning.

In summary, the main advantage of neural networks is that they can achieve good approximation accuracy with a reasonable number of parameters by training with data (hence, there is a lack of dependence on models). Therefore, neural network has been widely applied in many control areas such as fuel injection control, industrial robotics, and monitoring of mechanical machining process.

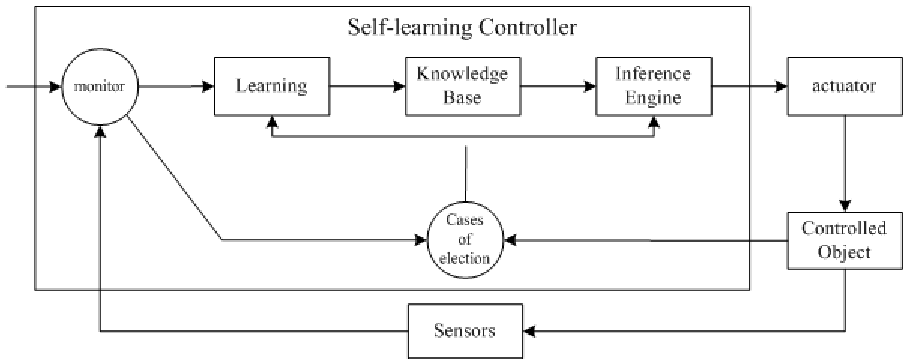
Learning Control Technology

Intelligent control is a control method mimicking human intelligence using computers. Developing a highly intelligent control system is one of the main research focuses, and the most important method in achieving this objective is learning control technology, alongside fuzzy logic, neural networks, evolutionary algorithms and expert systems. In an increasingly complex and ever-changing environment, the importance of learning control technology has been highlighted. Learning control is a kind of automatic control method that is able to learn from environmental information and unknown controlled plant information in the course of its operation, then after accumulating experience, it is capable of evaluating, classifying, making decisions and continuing to improve control performance. The main functions of learning control include searching, identification, memorizing, reasoning, modification and optimization. The problem of learning in automatic control systems has been studied in the past, especially in the late 1960s, and it has been the topic of numerous papers and books (Fu 1970, Mendel & Fu 1970, Sklansky 1966, Tsypkin 1971, Tsypkin 1968). Learning is a complex process which includes a wide range of learning content and a variety of ways. The ways of learning can be divided into mentor-learning, no-mentor learning and accentuation learning. Learning content can be divided into: learning control parameters - how to adjust controller parameters; learning controller structure - how to adjust the structure of the controller; learning environment – how to amend controlled activities when environment changes; learning uncertain complex controlled objects - how to simulate and approach plant model.

Learning is normally carried out in real time. It involves acquiring knowledge automatically and improves system control performance. The system has unknown characteristics that requires higher humanized learning functions. A typical learning algorithm without fixed structure is shown in Figure 3, the self-learning module is achieved by computer software. Learning control can make good use of historical information and experiences, and is adaptive to nonlinear dynamic system with uncertainties. However, it is not suitable for time-variant systems. The control strategy has been successfully used, for example, in the manufacturing sector, such as the control of an ultra-precision piezoelectric ceramic-driven tool, and tracking control system of the ellipse turning.

The main objective of learning control is to find a solution for system uncertainties and poor modeling. Considerable effort is required to reduce the difficulties of system control due to lack of necessary prior knowledge. It emphasizes memory, which is the main difference between learning and self-adaptive control. Learning control seeks the link between past experience

Figure 3. The structure of learning control system



and past control conditions, and adjusts to some appropriate controlled experience according to the situation.

The potentials of learning control in mimicking human behavior have shown tremendous superiority in the realization of intelligent control. There are currently three major research directions in the area including, 1) learning control based on pattern recognition, such as continuous learning, Bayes learning, stochastic approximation, and so on; 2) iterative learning control, mainly on periodicity operations of the nonlinear control systems, such as mechanical hand control of operations; and 3) connective learning control, mainly based on artificial neural network technology and evolutionary computation.

The development of learning control technologies is concentrated on the following areas including,1) learning control will develop from using a single technology to using a variety of technologies; 2) new learning algorithms will be developed for more complex tasks and can handle a variety of learning content 3) learning approach will change from a single approach to a hybrid direction of learning; and 4)the learning of control parameters will be changed to learning of changed structure, changed environment and complex unknown plants, and learning to control a variety of applications.

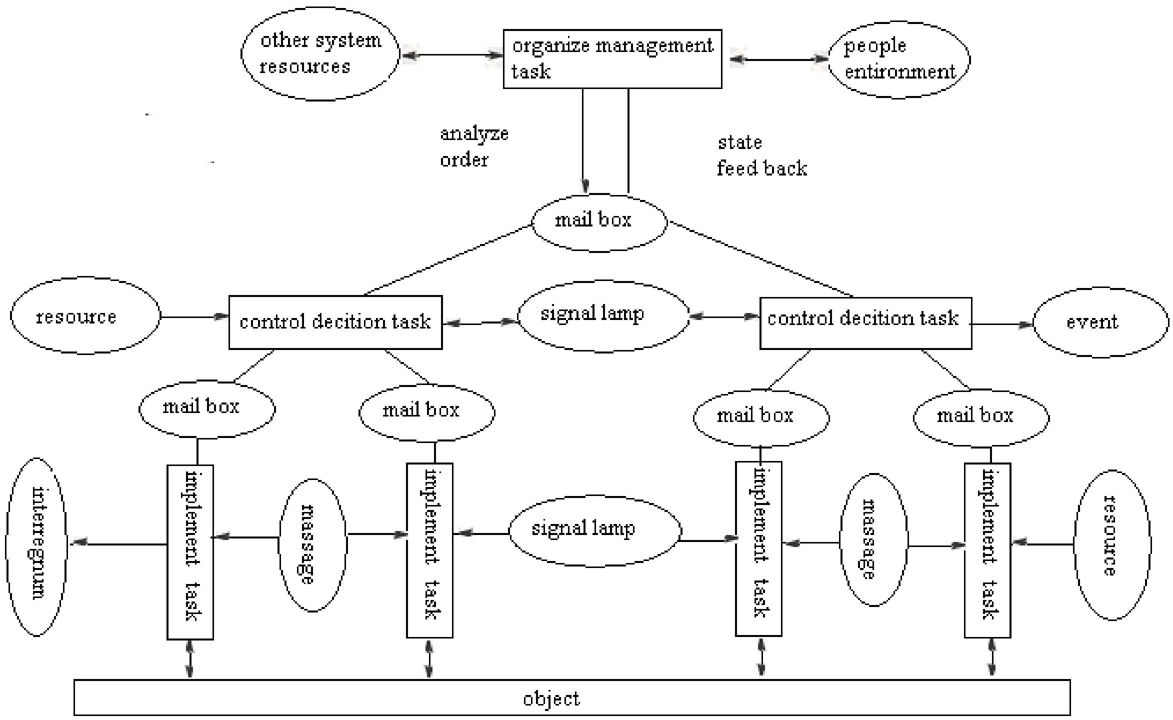
INTELLIGENT CONTROL SYSTEMS

Hierarchical Intelligent Control System

A typical intelligent control system uses a kind of hierarchical structure, which can be divided into an implementation level, a coordination level and an organization level. In accordance with Saridis's principle of "Increasing Precision and Decreasing Intelligence" (Saridis 1989) and the global optimization idea which covers from decision-making to control, the hierarchical intelligent control system can manage control tasks and processes hierarchically (Scattolini 2009). Therefore, the physical coordination of distributed subsystem can be implemented, and the goal of controlling the whole system can be achieved.

The implementation level of the hierarchical system can be realized by on-site control computer and the coordination level can be realized by manipulation stations and monitor computers. The organization level can be realized by management computers. The hardware structure of a hierarchical control system normally adopts network architecture of decentralized control and centralized management. Its logic structure adopts multi-task status table description structure of information transmission and mutual communication links, as shown in Figure 4.

Figure 4. The hierarchical structure of multi-task control system



Status Table

According to the task structure and interconnection mode shown in Figure 4, the hierarchical system divides a complex problem into several subsystems of different levels. Subsystems are operated at different rate in different time period.. The superior system modules hold the high-level command and distribute the decomposed tasks to the subordinate system modules. The subordinate system modules will return their running state to the superior systems. Consequently, the dynamic behavior of the system can be described by the state changes of each s module and task.

Both the status of tasks and the status change of the system can be fully indicated by the real-time status table. The status table updates synchronously in a form of input and output, and communicate with each other by sending messages through the link channel, and the communication channel has specified direction and the unique identifier, as the lights and mail boxes shown in Figure 4.

Each status table has limited number of statuses to describe a task, and the status transformation can be described as a series of optional commands or events triggered by order or priority. Here, the command can be the input, output or internal, such as a section of program or the action of external equipment. The executive time of command is tagged by a pair of time value.

Operation Object

In industrial control area, the widely-used real-time multi-task operating systems or real-time multi-task kernels usually regard operation object as task, event, interruption, resource, message, mailbox, signal lamp and so on. Interruption and task are considered as client objects and their tasks are continued by service objects that take communication responsibility. The task object

runs under the control of the scheduler and interruption objects respond to the interruption signals triggered by the physical environments (sensors, actuators, equipment, etc.). Service objects that provide client objects with simultaneous and coordinated action normally refer to signal (for a logo), message, mail box (a series of message queue sorted according to transfer protocol), events (broadcast news) and resources (equipments which can be shared and possess mutually exclusive function). The most basic operations of these objects include create, delete, suspend, resume, send message and so on.

Generally, there are two types of hierarchical model described by real-time status table: one is modeling of the environment and object, and the other is the description of system requirements the computer system. The superior and subordinate layers of the hierarchical system communicate through the information and control flow, and they also interact with outside equipments. A status table description structure uses a signal lamp to insert input, output and program segment into tasks, and uses mailbox to store the invoked command and program segment of asynchronous executive. Moreover, it adopts messages to provide direct communication between two status tables, and communication is required only when the two status tables are sending and waiting for messages during the same time interval.

Description of Structure

In a hierarchical real-time system, the main object that is able to denote action behavior is task (in Windows, it is known as the process and thread), so it is natural to denote the task object as the status table of description structure. However, other objects may also be denoted as status tables, because they also have activated action state, for example, resource object is used to show status table of display and access-read equipments, and interruption is used to show the status table of hardware equipment switch. Of course, we have to differ the given action behavior and the would-be action behavior, and to show the implementation results of monitoring different layers and command that the superior layer requests from the subordinate layer. The backstage object that manipulates and responds to environment is expressed as environment status table, and the objects that implement different commands in the front stage are expressed as system status table. Environment status table confirms its expression mode in accordance with action behavior and its connection with other status tables. For instance, if there is only one I/O command, and no matter whether or not there is internal command, it uses resource and / or interruption object to express environment status table according to the type of I/O command; if there are a serial of I/O commands, task objects are used to express the environment status table. However, in order to differ from the task object of environment status table, in accordance with the different I/O commands, it uses resource or interruption to connect all environment status tables and system status tables in the middle layers.

Generally speaking, system status table is expressed as the task of real-time system, and to use the name of status table as the identifier of task. In some cases, one task may produce several new tasks, so the system status table needs to be expressed as a parallel mechanism.

Expert Control System

Experts are referred to those people who possess profound professional knowledge and rich experience in solving specific problems. Expert system is mainly an intelligent computing system, which contains a great deal of professional knowledge and experience in a certain field and is able to cope with tough problems that can be only solved by experts in this field (Ligeza 1987). Therefore, an expert system contains tremendous professional knowledge and experience, and it combines artificial intelligent technology with computer

technology, and in accordance with knowledge and experience provided by experts, simulates the decision-making process of human experts. An advanced expert system even integrates knowledge and experience from many experts to solve more complicated problems.

Expert system possesses the following characteristics: The first is instructiveness. The structure of the problem one characteristic is going to solve is usually unreasonable, and its solution not only relies on theoretical knowledge and common sense, but also on the instructive knowledge of the expert system. The second is transparency. An expert system is able to explain its reasoning process and answer questions from system users, so as to ensure that users can understand this reasoning process and enhance users' trust on the expert system. The third is flexibility. This is referred to the ability to extend and enrich its knowledge, and to improve system performance, namely self-learning ability. The forth is symbol manipulation. Different from common data processing and numerical computing programs, an expert system emphasizes symbol processing and manipulation. It uses symbols to represent knowledge and symbol sets to represent the concept of problem. A symbol in expert system can represent a program design, and it can be used to represent any concept in the real world. The fifth is uncertain reasoning. The solutions mostly are experience-dependent, and generally experience knowledge is used to express uncertain problems, but there exists some probability. As many practical problems have uncertainties, their information presentation is usually not comprehensive, but expert systems can integrate all the information available, apply fuzzy theory and experience knowledge for reasoning.

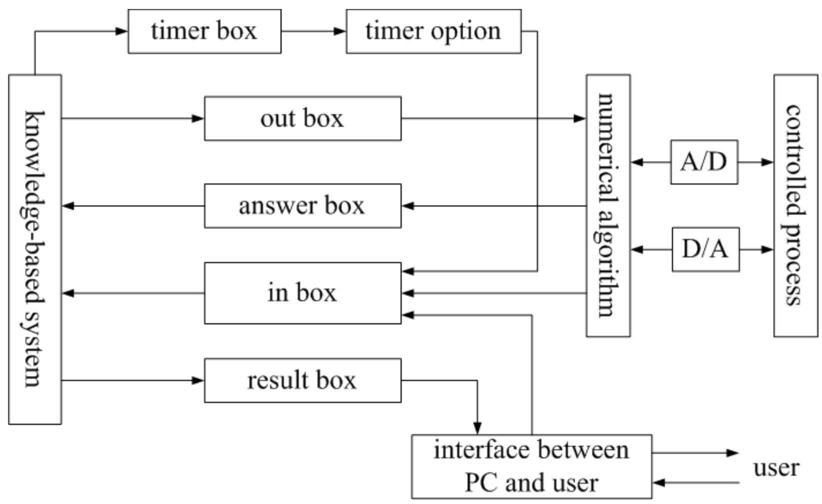
Expert systems can be divided as the following types based on their characteristics: explanation expert system, prediction expert system, diagnosis expert system, design expert system, planning expert system, monitoring expert system, controlling expert system, debugging expert system, teaching expert system and repairing expert system.

As shown in Figure 5, a typical expert control system usually contains two parts. Its numerical algorithm part is used for numerical computation, and contains quantitative analytical knowledge, and it connects directly with the controlled process. Knowledge-based system is used for symbolic reasoning, and contains qualitative instructive knowledge, and it is programmed according to design specification of expert system and connects with the controlled process through the numerical algorithm part. Numerical algorithm part is actually a knowledge base, which is made up of algorithms for control, judging and monitoring. The judging and monitoring algorithms compute control signal based on the control configuration commands and test signal. They extract feature information from numerical signal flow, and can be considered as a filter or a feature extractor. They send messages to the knowledge-based system only when system status experiences changes. In a stable state, the knowledge-based system does not function, and the system runs in the traditional control mode.

The communication among sub-processes is through the following five mailboxes:

- **Out box:** send control allocation command, parameter changes of control algorithms and information-send request from knowledge-based to numerical algorithm.
- **In box:** send the result of the algorithm, the inspection forecast information, the answer to information-send request, users' command and timing interruption signals respectively from the numerical algorithm, human-machine interface and timing operation system to knowledge-based system.
- **Answer box:** transmit numerical algorithm and communication answer signal for the information-send request of the knowledge-based system.

Figure 5. Structure diagram of expert controlling system



- **Result box:** transmit the results of human-machine communication from the knowledge-based system, including explanation of system operation situation of users' editing query for knowledge base, the reason for the algorithm implementation, reasoning condition and reasoning process tracking.
- **Timer box:** send the timing signal that the internal reasoning process of knowledge-based system needs to process an operation.

Expert control is one of the most active and widely used intelligent control methods in recent years. Engineers can design and apply different types of expert control systems for a variety of applications. For example, a general industrial expert controller can be designed for a simple control process (device), and an advanced expert control system is needed for a more complicated process control (device).

Neural Network Control System

Neural network control system is one kind of control system that uses neural network as the main development tool. In recent years, research on neural network control system has developed rapidly, and there are a large number of control structures and methods based on neural network, covering almost all research branches in the control areas (Ming & Mak 2000). Neural network control system can be divided into two types, which are inverse dynamic state control system and supervised control system.

Inverse Dynamic Control

When an inverse dynamic neural network model of a controlled plant is seemed as a controller and the controller is connected in series before the controlled plant, making up the inverse dynamic state control system.

Suppose the nonlinear dynamic equation of the controlled plant is $y(t)=F(y(t-1), u(t))$, $y(t)$ and $u(t)$ are the control output and input of the system, respectively. If $F(.)$ is reversible, the control signal $u(t)$ can be calculated by $y(t)$ and $y(t+1)$, that is:

$$u(t) = F^{-1}(y(t), y(t-1)) \tag{5}$$

$F^{-1}(\bullet)$is called the inverse dynamic equation of the controlled plant. A neural network can be trained to make the mapping relationship of input and output approach to $F^{-1}(\bullet)$, and this network is called the inverse dynamic neural network model of the controlled plant. As a controller, if the neural network model can fully approach the inverse dynamics of the controlled plant, the transfer function from the input u of controller to output y approximates to 1, that is the output y of the controlled object can follow the reference input of the controller, as a result, the control purpose can be achieved.

Generally, a neural network inverse dynamic control method contains two implementation stages: The first stage is the training period of the neural network controller, which is to study the inverse dynamics of the controlled plant with the help of neural network. Since neural network can approach any non-linear function, so in theory, as long as the object is reversible, its inverse dynamic model can be obtained with a neural network model. The inverse dynamic model is obtained by learning the input and output data of the controlled object, and accurate knowledge of the object is not needed. The second stage is the control operation stage, which is to connect the trained network controller with the system, and develop open-loop control for the controlled object. In normal circumstance, the output of the object would follow closely the input of the system.

The neural network inverse dynamic control structure is widely used in robot control, and there are many successful cases, but the main limitation is that it does not have control feedback. Therefore, it is necessary to adjust inverse dynamic model constantly through online learning.

Supervised Control

In practical applications, many control systems need people to participate, which mean that they need people to provide feedback control performance to complete specific tasks. If an automatic controller can be designed to simulate human's control role, this controller is called supervised control. In many cases, as it is difficult to get an analytical model of a plant, it is not esay to design a supervised control system. Besides methods of expert control or fuzzy control in the intelligent control filed, the neural network technology provides a new and effective method for supervised controller design.

Since neural network is superior in non-linear approximation and associative memory, it can be trained to approach the mapping relationship between human sensing input and decision-making output, so that a neural network controller can be obtained. The key to this method is that proper description data for control role of human is needed for the neural network to study and store.

Neural networks work as intelligent controllers in the following ways:

Direct Self-Tuning Control

A neural network first learns the inverse dynamic characteristics of the object off-line, which then works as feed-forward controller, and achieves adaptive control by using an online learning mechanism.

Indirect Self-Tuning Control

Using the self-learning function of neural network, when the controlled object parameters change, the change can be compensated by automatically adjusting controller parameters (e.g. PID parameters) and the system performance index can be guaranteed.

Neural Network Prediction Control

A neural network system works as a prediction model to provide forecast of a system, and maintains system performance by adjusting the system control input according to the forecast values.

Neural Network Model Reference Adaptive Control

This efficiently combines the Model Reference Adaptive Control (MRAC) of the linear system and neural network. A neural network model will be established and its weight will be adjusted based on the difference (y_m-y) between output y_m of reference model and the actual output y. Such a model will be then used for adaptive control.

Integrated Intelligent Control System

The core technologies of intelligent control include fuzzy control, expert system, neural network and evolutionary computation. Each of them has advantages, and disadvantages (Wang & Wang 2008). For example, the knowledge expression of the neural network is implicit and is difficult for people to understand. The ability to adapt to the outside world is limited; fuzzy set can describe complex or uncertain systems, but has no learning ability, hence difficult to amend the rules; expert system has the characteristics of knowledge transparence, which makes it easy to amend the rules, but requires computational power when there are too many rules, hence has limitations in real-time control. Therefore, it is necessary to integrate the intelligent control methods, and establish an information exchange and coordination mechanism among intelligent control methods and make use of complementary advantages for effective control. We have seen integrated approaches such as fuzzy neural network, fuzzy logic and expert system, fuzzy logic and evolutionary computation. So far, the widely used integration technology is the integration of fuzzy control and neural network technology.

There are different characteristics of fuzzy inference and neural network in the application of control systems. In general, fuzzy control is based on rules, and can be controlled well if there is enough knowledge on system control. While neural network needs a lot of data sample, and if the system has sufficient learning samples, neural network can achieve satisfactory control and continuous learning to correct its connection weights. Fuzzy mapping in the system follows the rules of mapping from set to set, but neural network is a point to point mapping. Therefore, fuzzy logic can easily express qualitative knowledge like human's control experience, but a neural network system has strong ability to learn by using quantitative data in the system.

The main weakness of fuzzy control is that the simple fuzzy process of information will lead to reduced control accuracy and dynamic performance. If we want to improve the precision, we have to increase quantization levels, which will expand the searching scope, increase computing work, and even lose the ability of real-time control; the design of fuzzy control is still lack of systematization, and the selection of rules and domain, the definition of fuzzy domain, and the selection of quantifiable factors, usually use the trial-and-error method. This is difficult for the control of a complex system.

The main problem in neural network control is that the performance of the neural network itself is not always optimal and has the local optimal problem, which leads to low efficiency of learning; In addition, the selection of neural network structure is still lack of theoretical basis, and neural network mapping theorem is the existence theorem, how to choose a neural network structure also takes the trial-and-error method.

From the characteristics and limitations of fuzzy logic and neural network control, we can see they are complementary. It is desirable to introduce a neural network structure to a fuzzy logic system to form a fuzzy neural network. The learning capability of the neural network can resolve the difficulty of identifying membership function and establishing rules, and realizing reasoning in the fuzzy logic system. but also can we give neural network the actual physical meaning, making it convenient for the control experience to join in. It is difficult for neural network to express the knowledge in control, which is the

basic function of fuzzy logic, and the capability of parallel learning of the neural network is just what fuzzy logic lacks when controlling a large and complex system, so the combination of the two can overcome the shortcomings of both.

Fuzzy neural network control technology is the essence of neural network integration control technology. Owing to limited learning ability for the fuzzy controller, and the lack of logic reasoning function for the neural network controller, soothe question is how to combine the two so that we can not only put our experience into the controller, but also make sure the controller can constantly learn and correct itself with the change of environment. This is an important step to achieve intelligent control. The combination of fuzzy control and neural network is the main direction of intelligent control technology development. The fuzzy neural network mainly has three structures, namely the input signal is the ordinary variables and connected weight is the fuzzy variable; the input signal is the fuzzy variable and the connected weight is the ordinary variable; the input signal and connected weight are both fuzzy variables.

Different fuzzy neural network models have different network topologies, and each structure has its own advantages. Comparing the similarities and differences of the structures of a network, something in common can be found. A Fuzzy neural network system basically has three-storey layers, which are fuzzy layer, fuzzy inference layer, and non-fuzzy layer.

The fuzzy layer completes the calculation of membership functions and calculates the membership degree of the variable in contrast to each fuzzy subspace. The fuzzy layer is a necessary part in all fuzzy neural networks.

The fuzzy inference layer connects with the premise and conclusion of fuzzy inference to realize fuzzy mapping. The structure of fuzzy inference is diversified - it can be a BP network, a RBF network, or other forms of network. different structure responds to different algorithm, which is just the difference among the various models.

The de-fuzzy layer changes distributed fuzzy state of reasoning conclusion variable into a determined state, and is responsible for working out a certain output. The common methods for the de-fuzzy layer are maximum membership principle and fuzzy cancroids. In some specific networks, there is no need to construct a de-fuzzy layer.

In order to enhance the adaptability of a fuzzy neural network, the fuzzy layer, fuzzy inference layer and de-fuzzy layer are usually made up of multi-layer networks. In this way, the automatic adjustment of membership function and fuzzy regulation in fuzzy inference can be achieved by learning through the Internet.

Network-Based Intelligent Control System

With the popularity of the internet, the Open System Interconnect Reference Model (OSI) protocol has been replaced by internet protocol, which has been approved by International Standards Organization (ISO) and become a practical network standard. The IP is not only supported by the Ethernet, but also supported by other networks. Therefore, the integration of information network and control network has become the direction of enterprise integration automation, which is mainly demonstrated in the following three aspects,1) the integration of Control Network and Information Network: On the one hand industrial process control, no matter it is field bus level or distributed control system, even though we can realize "accurate control" according to the given object parameter value, but whether or not it is optimal, reasonable and it should be to continue, the answer is the fact that process-level computers also need management information system (MIS) to give parameters; On the other hand, although the traditional MIS (the original office automation management information system), has the support of excellent human-resource management, financial management, quality management, equipment management software, as the data has

to be input with manual work, which is labor and time-consuming, incorrect as well, so it is difficult to play a role. Therefore, MIS and industrial process control must be integrated, otherwise MIS in the enterprise will not be of much practical significance. Experience has proven that not only does MIS need to integrate with the process, but also that the process control system (control network) needs to integrate with MIS.

As Internet is the largest computer network in the world, which transmits information in high speed with low cost, it provides a mechanism for enterprises to connect their MIS with the Internet. The Internet not only expands the information channels between the enterprises and external sources, but also penetrates into the enterprises, becoming a brand new internet- Intranet. It provides a changing, open, easy-to-use two-way multimedia information exchange environment instead of the traditional MIS, which is a static, closed, complex text digital environment. Thus, the MIS of enterprises have experienced revolutionary change. As the Intranet is integrated into traditional MIS, full use can be made of existing MIS platform system resources, and with the reduction in construction cost, higher efficiency can be achieved.

On-site Control, MIS integration with the Internet: The use of the field bus control, MIS and Internet are now possible, and the technologies are in place. The MIS plays an important role if attempting to integrate the three technologies. This integration will inject new vitality into the enterprises, and achieve comprehensive and integrated automation in enterprises, enterprise groups and the whole field. Enterprises can communicate with other enterprises, and also with international counterparts.

The integration can be delivered using the bus network, Local Area Network (LAN), and the Internet. The first layer is the control network, which is made up of on-site control equipment and is used to control and detect equipment. The second layer is the process monitoring control. The workstation, PC or controller, which takes the responsibility of monitoring, usually works as LAN nodes. The on-site bus network connects with the process monitoring layer through a special communication interface. The monitoring layer performs the tasks of online monitoring, control system and error diagnosis. The third layer is the enterprise management layer, which is normally called the information layer. This layer is for achieving enterprise information integration and management, and control integration. Its network nodes are mainly made up of high-performance computers, workstations, PC. The enterprise management layer can achieve information integration with remote network nodes through the Internet.

In the three-layer structure, an OPC (OLE for Process Control) server plays an important role. The OPC server provides a standard interface to the upper server applications, enabling upper server applications, visiting the data in servers and realizing the connection with upper applications. Real-time data from equipment is provided through the OPC to data invokers. The OPC server is the basis of information integration, which enables the sharing of information in distributed control applications.

While the intelligent distributed control system is for the integration of intelligent devices, it also needs to make effective use of all intellectual activities, and through knowledge bases, databases, computer and communication network, it is able to integrate activities such as, order, designing, production and sales, and forms production system to improve the overall efficiency. From the perspective of the entire control process, order, designing, production and sales are independent in function, and their solutions to the problems are different, and while each function subsystem finish its processing sub-task, they also coordinate mutually with each other, and in a manufacturing enterprise, the entire manufacturing process from designing to sales is a typical multi- autonomous agent solution process. Collaborative technology

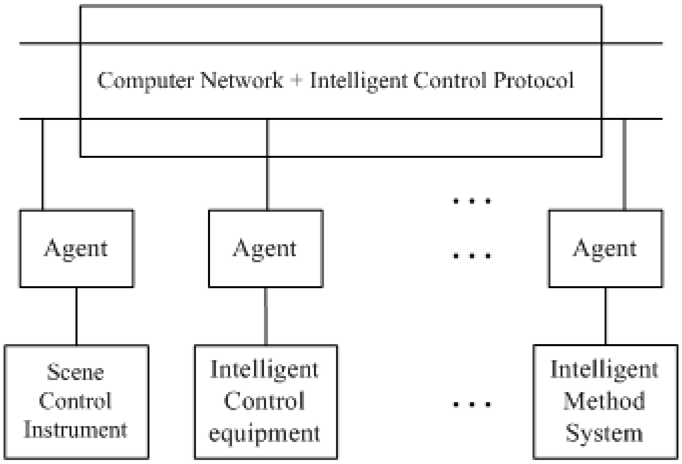
is a newly-developed issue in recent years, and the relevant discipline is often called Computer Supported Collaborative Work (CSCW) or groupware. Groupware is the system that provides members with support by using computers and communication network, and makes them coordinate well. The related research of groupware on the theory, technology, method and application belong to the area of CSCW research. Coordination is not only an important characteristic in the application of CSCW, but also an important distinction between the application of CSCW and traditional computer application. So coordination is the core issue of CSCW.

In the field of artificial intelligence, agent technology can be used to denote a certain intelligent entity, so as to distinguish it from general entity. They can be either physical entities or abstract entities, and can be decision-making, reasoning and problem solving - a kind of intelligent logic unit. An agent-based model is a personification model, and can unify the human behavior in the control system and others of the control unit, to provide a unified description of methods. In artificial distributed intelligence, agents, as smart nodes on the network, connect with each other through the computer network, form a distributed Multi-Agent System (MAS). Distributed artificial intelligence has two branches, which are distributed problem solving (DPS) and MAS. MAS have the ability to self-organize dynamically in respond to environmental changes. The area has gained much attention in many areas. In real world, as knowledge and activities are distributed in nature, Distributed Artificial Intelligence (DAI) needs to be investigated. DAI possesses the characteristics of high-efficiency, high-reliability and low-cost, which make it possible industrial applications. DAI also provides an approach to study open systems. With the application of DAI theory in various fields, the meaning of agent has been expanded. In some circumstances, agent denotes entity that has a closed function and decision-making ability, known as " autonomous agent "; In other cases, agents play a role as the original meaning, only deals with external affairs on behalf of function entities, namely according to the operation situation of function entities (the parties) it is responsible for external cooperation and coordination.. Whether it is a "self-determination" or "agent", the agent connects through the network and works by transmitting messages to each other. As the agent has the characteristic of personification, the language of Agent is close to natural language. In order to improve the adaptability of manufacturing systems, the design of distributed intelligent control system must follow the principle of openness. As indicated by Hewitt, an open information system refers to a system that one can never expect its operating results and at any time it can receive external information. The concrete manifestation of "Openness" is as follows: The openness of the tasks: tasks can be input at any time and dealt with immediately; The openness of the system: it can accept the changes from the system inside (such as failure), and accept outside interference on the system (such as changes in physical device configuration); The openness of solving the problem, the process of problem-solving receives changes of information and knowledge.

To achieve the three openness mentioned above, intelligent distributed control system uses distributed architecture, giving the components of entities or sub-systems greater autonomy to form intelligent autonomous agents, and connects with computer networks in the form of intelligent nodes, and all nodes are equal on the logic (there is no direct control relationship between each other), dispersed in physical locations, and independent in function, there exists a loose coupling relationship among each nodes. Nodes communicate by sending messages and realize mutual coordination on the basis of common communication language. The Distributed Control System structure is shown in Figure 6.

The rapid development of intelligent distributed control changes the narrow understanding of intelligent control systems. Attention has been

Figure 6. Intelligent distributed manufacturing structure based on computer network



placed on the man-machine integration and intelligence complementation. The integration road of Intelligent Control has not unified and standard Judging rules to follow. From the current situation, it is difficult to form a general, unified theory for intelligent distributed control, but the establishment of an integrated intelligent control framework is quite realistic and necessary. The framework should be open, form-diverse, and within which people can work as system members.

CHALLENGES OF INTELLIGENT CONTROL TECHNOLOGIES

With the rapid development of information technology, computer science and technology, biomedical technology, and automatic technology, the intelligent control is faced with many challenges, and they are described in the following.

Design support tools need to be developed to automatically generate alternative solutions (models of processes and controllers). Some solutions could be deployed simultaneously into self-organizing controllers, by using evolutionary programming as optimization techniques in an integrated optimal design of process and control strategies.

Biological organisms are equipped with highly efficient, redundant systems for sensing the environment, processing and storing the information acquired. For advanced control systems, such as man-made systems, we can achieve major progress by using much more affordable sensor technologies and by developing new tools to interpret the data acquired and to represent them in the form of knowledge, thus adding truly cognitive functions. The adaptation of control laws, or the complete dynamic reconfiguration of control strategies must be included into the new generation of control systems. The general design philosophy must shift from resource limitation to resource adequacy rendering more understandable solutions possible.

The new generation of control systems should be able to autonomously control complex, poorly understood processes, such that some well-designed goal can be achieved. The intelligent control systems should cope with changes not anticipated in the process or its environment, learn from past experience, actively acquire and organize knowledge about the surrounding world and plan its future behavior.

The incorporation of intelligent techniques (e.g. fuzzy logic, neural networks, and genetic algorithms) into advanced control systems, by

employing alternative representation schemes and formal methods to add additional relevant information that cannot be used in the standard control-theoretical framework of differential and difference equations, represents a new challenge for the control systems community.

Fuzzy logic systems are suitable to represent qualitative knowledge, either provided by human experts, or automatically acquired from data (rule induction, learning). Neural networks can perform complex learning and adaptation tasks by imitating the function of biological neural systems, and thus can be used as models for nonlinear, multivariable systems, trained by using input-output data observed on the system.

Recent developments in computer networks and communications, combined with new ways of information and knowledge processing, provide new possibilities for control purposes in distributed environment.

The communication network represents a most important element of assure the stability and performances of remote control systems, for which new concepts of distributed control, real-time communication networks, and autonomous control strategies must be developed. The orientation towards a network-centered view has different architectures of computer control systems, enabling various forms of telepresence.

An important area of application of control theory and engineering is the control of communication networks including variable time delays. A real issue for control engineering is the design and application of instruments and control schemes for telemonitoring, telepresence, and telecontrol.

A new generation of control systems has to be developed by increasing the autonomy of control systems and their degree of intelligence. Through the integration of model-based and intelligent methodologies, it is possible to solve the robustness problems for complex systems with large uncertainties. Hybridization of intelligent methodologies by transformation, combination, fusion, and association will increase the autonomy of control systems, while the combination of these hybrid technologies with agent technology could offer valuable solutions for the modeling and control of complex systems.

Hybrid technologies have to be used to integrate the mathematical models of discrete event and continuous systems with the intelligent techniques and intelligent agents to model and control the complexity. The most important attributes of the intelligent systems, such as perception, communication, learning, planning, and behavior generation, reasoning and thinking, are more and more included in different forms in the advanced autonomous control systems.

The integration of intelligent agents into hierarchical and heterarchical architectures with different time scale and resolution allows for the development of a new generation of control systems organized on the "Increasing Precision and Decreasing Intelligence" principle, as a multiresolutional system.

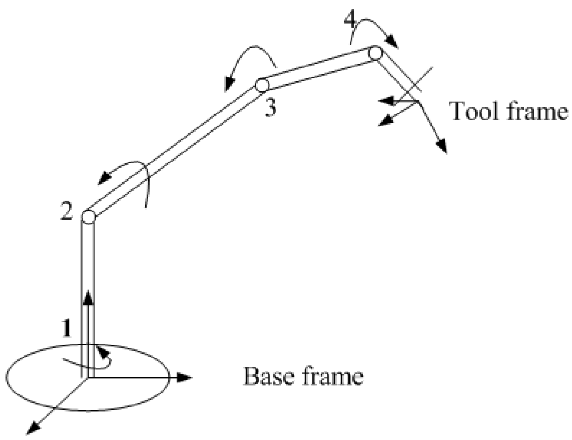
NEURAL NETWORK BASED ROBOT CONTROL: A CASE STUDY

An important area of application of neural networks is in the field of robotics. Usually, neural network based methods are designed to control a manipulator, which is the most important form of industrial robots, for example, to grasp objects, based on sensor data. Another application includes the steering and path-planning of autonomous robotic vehicles. In robotics, the major task involves controlling position based on sensor feedback. There are normally four problems in robot control, and they are described as follows:

Forward Kinematics

Kinematics is the science of motion which treats motion without considering forces applied. Within this science one studies the position, velocity, acceleration, and all higher order derivatives of the

Figure 7. The structure of robot manipulator



position variables. A very basic problem in the study of mechanical manipulation is the forward kinematics. This is the static geometrical problem of computing the position and orientation of the end-effector ('hand') of a robot manipulator. Specifically, given a set of joint angles, the forward kinematic problem is to compute the position and orientation of the tool frame relative to the base frame (as shown in Figure 7).

Inverse Kinematics

This problem is posed as follows: given the position and orientation of the end-effector of the manipulator, calculate all possible sets of joint angles which could be used to attain this given position and orientation. This is a fundamental problem for controlling manipulators. The inverse kinematic problem is not as simple as the forward one for serial robots. Because the kinematic equations are nonlinear, their solution is not always easy or even possible in a closed form. The questions of the existence of a solution or multiple solutions will arise. This problem has to be solved for most of robot control systems.

Dynamics

Dynamics is a field of study devoted to studying the forces required to cause motion. In order to accelerate a manipulator from rest, maintain a constant end-effector velocity, and finally decelerate to a stop, a complex set of torque outputs must be applied by the joint actuators. In dynamics not only the geometrical properties (kinematics) are used, but also the physical properties of the robot are taken into account. For instance, the weight (inertia) of a robot arm determines the force required to change the motion of the arm. The dynamics introduces two extra problems to the kinematic problem.

1. The robot arm has a 'memory'. It's respond to a control signal depends on its history (e.g. previous positions, speed, acceleration).
2. If a robot grabs an object, its dynamics changes but the kinematics doesn't. This is because the weight of the object has to be added to the weight of the arm.

Trajectory Generation

To move a manipulator from one position to another in a smooth and controlled fashion, each joint must be moved accordingly and this is determined by a defined trajectory. To compute the motion functions of each joint is the problem of trajectory generation. The following sections of this chapter will discuss the problems associated with the positioning of an end-effector.

End-Effector Positioning

The fgoal in robot manipulator control is often the positioning of an end-effector in order to be able to, e.g., pick up an object. To control a robot arm accurately, the following steps are involved:

1. Determine the target coordinates related to the base of the robot. Typically, when this

position is not always the same, this is done with a number of fixed cameras observing the work scene, from the image frame determine the position of the object in that frame, and perform a pre-determined coordinate transformation.
2. With a precise model of the robot (supplied by the manufacturer), calculate the joint angles to reach the target (i.e., the inverse kinematics).
3. Move the arm (dynamics control) to the position.

If these parts are relatively simple to solve with a high accuracy, why involve neural networks? The reason is the applicability of robots. When 'traditional' methods are used to control a robot arm, accurate models of the sensors and manipulators (in some cases with unknown parameters which have to be estimated from the system's behavior; yet still with accurate models as starting point) are required and the system must be calibrated. Also, systems which suffer from wear-and-tear (and which mechanical systems don't?) need frequent recalibration or parameter determination. Finally, the development of more complex (adaptive!) control methods allows the design and use of more flexible (i.e., less rigid) robot systems, both on the sensory and motory side.

Camera-Robot Coordination

The system we focus on in this section is a work OOR observed by fixed cameras, and a robot arm. The vision system must identify the target as well as determine the visual position of the end-effector.

The target position X_{target} together with the visual position of the hand X_{hand} are input to the neural controller *N*(.). This controller then generates a joint position θ for the robot:

$$\theta = N(X_{target}, X_{hand}) \tag{6}$$

We can compare the generated θ with the optimal θ_0 generated by a fictitious perfect controller *R(.)*:

$$\theta_0 = R(X_{target}, X_{hand}) \tag{7}$$

The task of learning is to make it possible for the function *N* to generate an output close enough to θ_0. There are two problems associated with training *N*(.):

1. Generating learning samples which are based on equation 7. This is not trivial, as *R(.)* is an unknown function. Self-supervised or unsupervised learning is required. Some examples to solve this problem are given below.
2. Constructing the mapping *N(.)* from the available learning samples. When the learning samples are available, a neural network uses these samples to represent the whole input space over which the robot is active. This is evidently a form of interpolation, but has the problem that the input space is of a high dimensionality, and the samples are randomly distributed.

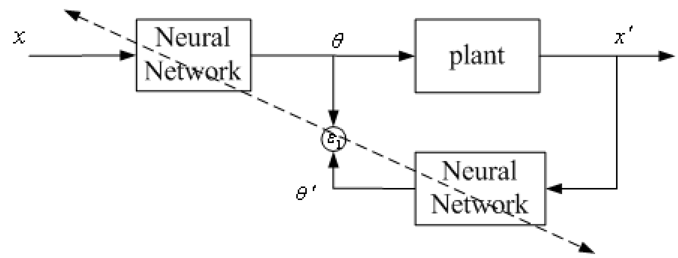
Implementation-Feed-Forward Networks

When a feed-forward system is used for controlling the manipulator, a self-supervised learning system is required. One such a system has been reported by Psaltis, Sideris and Yamamura (Psaltis, Sideris, & Yamamura, 1988). Here, the network is used to control the two-dimensional positioning of a robot arm.. Three methods are developed including

Indirect Learning

As shown in Figure 8, in indirect learning, a Cartesian target point *x* in world coordinates is generated, e.g., inspecting an object using two cameras. This target point is fed into the network, which

Figure 8. Indirect learning system for robotics



generates an angular vector θ. The manipulator then moves to the position θ, and the cameras determine the new position x' of the end-effector in world coordinates. This x' again is input to the network, resulting in θ'. The network is then trained to reduce the error $\varepsilon_1 = \theta - \theta'$

However, minimisation of ε_1 does not guarantee minimisation of the overall error $\varepsilon = x - x'$. For example, the network often settles at a 'solution' that maps all x's to a single θ (i.e., the mapping).

General Learning

The method is very much similar to supervised learning, but the input θ must be provided by a user. Thus the network can directly minimize $|\theta - \theta'|$. The success of this method depends on the interpolation capabilities of the network. Correct choice of θ may pose a problem.

Specialized Learning

The goal of the training of the network is to minimise the error at the output of the plant: $\varepsilon = x - x'$. We can also train the neural network by "back-propagating" this error through the plant (compare this with the back-propagation of the error). This method requires knowledge of the Jacobian matrix of the robot. A Jacobian matrix of a multidimensional function F is a matrix of partial derivatives of F, i.e., the multidimensional form of the derivative. For example, if we have $Y = F(X)$, i.e.,

$$\begin{aligned} y_1 &= f_1(x_1, x_2 \ldots x_n), \\ y_2 &= f_2(x_1, x_2 \ldots x_n), \\ &. \\ &. \\ &. \\ y_m &= f_m(x_1, x_2 \ldots x_n) \end{aligned} \tag{8}$$

then

$$\begin{aligned} \delta y_1 &= \frac{\partial f_1}{\partial x_1}\delta x_1 + \frac{\partial f_1}{\partial x_2}\delta x_2 + \ldots + \frac{\partial f_1}{\partial x_n}\delta x_n, \\ \delta y_2 &= \frac{\partial f_2}{\partial x_1}\delta x_1 + \frac{\partial f_2}{\partial x_2}\delta x_2 + \ldots + \frac{\partial f_2}{\partial x_n}\delta x_n, \\ &. \\ &. \\ &. \\ \delta y_m &= \frac{\partial f_m}{\partial x_1}\delta x_1 + \frac{\partial f_m}{\partial x_2}\delta x_2 + \ldots + \frac{\partial f_m}{\partial x_n}\delta x_n \end{aligned} \tag{9}$$

or

$$\delta Y = \frac{\partial F}{\partial X}\delta X \tag{10}$$

can be written as

$$\delta Y = J(X)\delta X \tag{11}$$

where J is the Jacobian matrix of F. The Jacobian matrix can be used to calculate the change in the function when its parameters change.

We have

$$J_{ij} = \left[\frac{\partial P_i}{\partial \theta_j}\right] \tag{12}$$

where $P_i(\theta)$ the i^{th} element of the plant output for input θ. The learning rule applied here regards the plant as an additional and unmodiable layer in the neural network. The total error $\varepsilon = x - x'$ is propagated back through the plant by calculating the δ as in equation (13).

$$\delta_j = F'(s_j)\sum_i \delta_i \frac{\partial P_i(\theta)}{\partial \theta_j},$$
$$\delta_i = x_i - x_i', \tag{13}$$

where i iterates over the outputs of the plant. When the plant is an unknown function, and $\frac{\partial P_i(\theta)}{\partial \theta_j}$ can be approximated by

$$\frac{\partial P_i(\theta)}{\partial \theta_j} \approx \frac{P_i(\theta + h\theta_j e_j) - P_i(\theta)}{h} \tag{14}$$

where e_j is used to change the scalar θ_j into a vector. This approximation derivative can be measured by slightly changing the input to the plant and measuring the changes in the output.

A two-layer feed-forward network is trained with back-propagation. However, instead of calculating a desired output vector given an input vector which should have invoked the current output vector is reconstructed, and back-propagation is applied to this new input vector and the output vector. The configuration used consists of a monocular manipulator which has to grasp objects. Due to the fact that the camera is fixed to the robot, the task is to move the hand such that the object is in the centre of the image and has some predetermined size. The process consists of the following operations:

1. Measure the distance from the current position to the target position in camera domain, x.
2. Use this distance, together with the current state θ of the robot, as input for the neural network. The network then generates a joint displacement vector $\Delta\theta$.
3. Send $\Delta\theta$ to the manipulator.
4. Again measure the distance from the current position to the target position in camera domain, x'.
5. Calculate the move made by the manipulator in visual domain, $x - {}_t^{t+1}Rx'$ where ${}_t^{t+1}R$ is the rotation matrix of the second camera image with respect to the r^{st} camera image.
6. Train the neural network using the learning pair $(x - {}_t^{t+1}Rx', \theta; \Delta\theta)$.

Through experiments, it is found that this system is able to learn the defined correct behaviors in only tens of iterations, and is adaptive to changes in the sensor or manipulator.

By using a feed-forward network, the available learning samples are approximated by a single, smooth function consisting of a summation of sigmoid functions. As mentioned in the previous section of the chapter, a feed-forward network with one layer of sigmoid units is capable of representing any function. But how are the optimal weights determined within a finite time to obtain this optimal representation? Experiments have shown that, although a reasonable representation can be obtained in a short period of time, an accurate representation of the function that governs the learning samples is often not feasible or extremely difficult. This is due to the global characteristics of the approximation obtained with a feed-forward network with sigmoid units. Every weight in the network has a global effect on the

final approximation that is obtained. Building local representations is an obvious way out as every part of the network is responsible for a small subspace of the total input space. Thus accuracy is obtained locally. This is typically obtained with a Kohonen neural network.

SUMMARY

Intelligent control achieves automation via the emulation of biological intelligence. It either seeks to replace a human for performing a control task (e.g., a chemical process operator) or it borrows ideas from how biological systems solve problems and applies them to control problems (e.g., the use of neural networks for control).

In this chapter, we first gave a brief overview of the foundations of intelligent control. This includes the difference between conventional and intelligent control, the concepts of intelligence and intelligent control, the definition of intelligent system, and the notions of control and intelligent system. The models for intelligent controllers such as the DES model and the hybrid system model were presented, and the relative intelligent control technologies such as fuzzy control, neural network control and learning control were discussed. The basic types of intelligent control systems including hierarchical control system, expert control system, neural control system, integrated intelligent control system, and network based intelligent control system were investigated. The challenges of intelligent control were discussed and a case study was presented. An intelligent manufacturing system is proposed to demonstrate that the application of intelligent control theory and technology into the manufacturing area is able to improve the manufacturing processes.

With the development of intelligent technology such as information technology, computer technology, automation technology, artificial intelligence and life science, intelligent control faces many challenges and gains lots of promising development opportunities. The main areas of interest in the control community in the future include agent technology, architecture-based design, artificial intelligence, concurrent engineering, compensability, design patterns, distributed embedded systems, model-based software engineering, modular systems, object-oriented programming, product line engineering, real-time distributed systems, reusability, software components and software processes. The study of intelligent control is attracting more and more interest. It is predictable that intelligent control will have broader and broader applications and will offer useful solutions to the challenging problems in the future.

REFERENCES

Agogino, A., Stanley, K., & Miikkulainen, R. (2000). Online Interactive Neuro-evolution. *Neural Processing Letters*, *11*(1), 29–37. doi:10.1023/A:1009615730125

Albus, J. S. (1991). Outline for a Theory of Intelligence. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*(3), 432–509. doi:10.1109/21.97471

Alur, R., Courcoubetis, C., Halbwachs, N., Henzinger, T. A., Ho, P. H., & Nicollin, X. (1995). The algorithmic analysis of hybrid systems. *Theoretical Computer Science*, *138*(1), 3–34. doi:10.1016/0304-3975(94)00202-T

Alur, R., & Dill, D. (1994). The theory of timed automata. *Theoretical Computer Science*, *126*(2), 183–235. doi:10.1016/0304-3975(94)90010-8

Antsaklis, P.J. (1994). Defining intelligent control. *IEEE Control Systems*, (pp. 4-5 & 58-66). Report of the Task Force on Intelligent Control, P.J Antsaklis, (Chair).

Antsaklis, P. J. (1997). Intelligent control . In *Encyclopedia of Electrical and Electronics Engineering*. Hoboken, NJ: John Wiley & Sons, Inc.

Antsaklis, P. J., Lemmon, M., & Stiver, J. A. (1996). Learning to be autonomous: Intelligent supervisory control . In Gupta, M. M., & Sinha, N. K. (Eds.), *Intelligent Control Systems: Theory and Applications* (pp. 28–62). Washington, DC: IEEE Press.

Antsaklis, P. J., & Nerode, A. (1998). Hybrid control systems: An introductory discussion to the special issue. In P.J. Antsaklis & A. Nerode, (eds.), IEEE Trans. Automatic Control, 43(4), 457-460.

Antsaklis, P. J., & Passino, K. M. (1993). Introduction to intelligent control systems with high degrees of autonomy . In Antsaklis, P. J., & Passino, K. M. (Eds.), *An Introduction to Intelligent and Autonomous Control* (pp. 1–26). Amsterdam: Kluwer.

Branicky, M. S. (1995). Universal computation and other capabilities of hybrid and continuous dynamical systems. *Theoretical Computer Science*, *138*(1), 67–100. doi:10.1016/0304-3975(94)00147-B

Celano, G., Costa, A., & Fichera, S. (2000). Fuzzy Scheduling of A Flexible Assembly Line Through An Evolutionary Algorithm. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, (pp. 328-333).

Chen, W. W., Wei, X. P., & Zhao, T. T. (2008). *Product schemes evaluation method based on improved BP neural network*, (. *LNCS*, *5227*, 99–106.

Fu, K. S. (1986). Learning Control Systems--Review and Outlook. *IEEE Transactions on Automatic Control*, *PAMI-8*(15), 327–342.

Koutsoukos, X. D., He, K. X., Lemmon, M. D., & Antsaklis, P. J. (1998). Timed Petri nets in hybrid systems: Stability and supervisory control. *Journal of Discrete Event Dynamic Systems: Theory and Applications*, *8*(2), 137–173. doi:10.1023/A:1008293802713

Lei, L., Wang, H. J., & Cheng, Y. F. (2006). *An improved BP algorithm based on global revision factor and its application to PID control*. (. *LNCS*, *3972*, 1019–1026.

Lemmon, M. D., He, K. X., & Bett, C. J. (1998). Modeling hybrid control systems using programmable Petri nets. In *3rd International Conference ADMP'98, Automation of Mixed Processes: Dynamic Hybrid Systems*, (pp 177-184).

Ligeza, A. (1987). Expert systems approach to decision support. *European Journal of Operational Research*, *37*(1), 100–110. doi:10.1016/0377-2217(88)90284-6

Mak, K. L., & Wong, Y. S. (2000). Genetic Design of Cellular Manufacturing Systems. *Human Factors and Ergonomics in Manufacturing*, *10*(2), 177–192. doi:10.1002/(SICI)1520-6564(200021)10:2<177::AID-HFM4>3.0.CO;2-A

Mendel, J. M., & Fu, K. S. (1970). *Adaptive, Learning and Pattern Recognition Systems*. New York: Academic Press.

Ming, X. G., & Mak, K. L. (2000). Intelligent setup planning in manufacturing by neural networks based approach. *Journal of Intelligent Manufacturing*, *11*(3), 311–331. doi:10.1023/A:1008975426914

Moody, J. O. (1997). *Petri Net Supervisors for Discrete Event Systems*. PhD thesis, Department of Electrical Engineering, University of Notre Dame, Notre Dame.

Moody, J. O., & Antsaklis, P. J. (1998). *Supervisory Control of Discrete Event Systems using Petri Nets*. Kluwer Academic Publishers.

Passino, K. M., & Antsaklis, P. J. (1993). Modeling and analysis of artificially planning systems . In Antsaklis, P. J., & Passino, K. M. (Eds.), *An Introduction to Intelligent and Autonomous Control* (pp. 191–214). Amsterdam: Kluwer Academic Publishers.

Passino, K. M., & Lunardhi, A. D. (1996). Qualitative analysis of expert control systems . In Gupta, M. M., & Sinha, N. K. (Eds.), *Intelligent Control Systems: Theory and Applications* (pp. 404–442). Washington, DC: IEEE Press.

Petridis, V., & Kaburlasos, V. G. (1998). Fuzzy lattice neural network (FLNN): a hybrid model for learning. *IEEE Transactions on Neural Networks*, *9*(5), 877–890. doi:10.1109/72.712161

Psaltis, D., Sideris, A., & Yamamura, A. A. (1988). A multilayered neural network controller. *IEEE Control Systems Magazine*, *9*(3), 17–21. doi:10.1109/37.1868

Ramadge, P. J., & Wonham, W. M. (1989). The control of discrete event systems. *Proceedings of the IEEE*, *77*(1), 81–89. doi:10.1109/5.21072

Saridis, G. N. (1989). Analytic formulation of the principle of increasing precision with decreasing intelligence for intelligent machines. *IFAC Proceedings Series*, (10), 529-534.

Scattolini, R. (2009). Architectures for distributed and hierarchical Model Predictive Control - A review. *Journal of Process Control*, *19*(5), 723–731. doi:10.1016/j.jprocont.2009.02.003

Sklansky, J. (1966). Learning Systems for Automatic Control. *IEEE Transactions on Automatic Control*, *11*(1), 6–19. doi:10.1109/TAC.1966.1098229

Tsypkin, Y. A. Z. (1968). Self-Learning-What is it? *IEEE Transactions on Automatic Control*, *AG-13*(6), 608–612. doi:10.1109/TAC.1968.1099015

Tsypkin, Y. A. Z. (1971). *Adaptation and Learning in Automatic Systems*. New York: Academic Press.

Wang, W., & Wang, H. L. (2008). Study on integrated intelligent control with hierarchical structure. In *Proc of International Symposium on Computer Science and Computational Technology*, (Vol. 2, pp. 63-65).

White, D. A., & Sofge, D. A. (1992). *Handbook of Intelligent Control Neural, Fuzzy, and Adaptive Approaches*. New York: Van Nostrand Reinhold.

Zeng, Y. J., Wang, X. D., & Wang, S. J. (1998). Improved BP algorithm for pattern recognition. In *Proc. International Conference on Signal Processing*, (Vol. 2, pp. 1327-1330).

Chapter 10
Intelligent Product Design:
Intelligent CAD

CAD (Computer Aided Design) is almost instead of classical designing method which drawing plan on paper nowadays. With the development of information technology, the traditional CAD technology becomes rather matured and is developing towards a modern direction of being further integrative, intelligent, and collaborative, namely ICAD (Intelligent CAD). ICAD is a complex system consist of multi agents or multi experts to design product. It can simulate expert in this area to help designer accomplish design. ICAD is based on some technology such as artificial intelligence, CAD technique, expert systems technique, modern mechanical design theory and database technique. In this chapter, the reason of ICAD proposing is given firstly, then some research and application is described on the second sector. Thirdly, some theory and technique about ICAD is discussed. Finally, a case study is presented.

DOI: 10.4018/978-1-60566-864-2.ch010

INTRODUCTION

Inevitability of ICAD

Design is an important activity in human's realization and reform of the world, and all the creativity by human beings derive from the design. The aim of the product's design is to make sure the function of the product and establish a system of good function, low cost, proper price, etc, which play a decisive role in the technology and beneficial result of the product. Especially in the market economy environment, design is the key to win, and it plays an important role in the development of national economy.

From globalization to mass customization, several trends are affecting the development of CAD and manufacturing tools. Globalization is exerting an influence on the development of CAD. Companies' running, competition and markets have

gone global, and in response, CAD is taking on more functionality and allowing for increasing collaboration demand between departments.

The major inevitability of ICAD is discussed as follows.

Globalization

Sending jobs to foreign locations continues to be a prevalent and highly controversial-trend. It's driven by such factors as increasing globalization, the ubiquity of high-speed Internet connections and the rising pressure to increase corporate profits. In fact, in the CAD world, most developers have already placed software teams in India and/or Russia to handle research, development and support. And along with manufacturing, data processing and call center jobs, companies are also taking high-tech engineering design and development work. For example, multidiscipline engineering design and drafting, software development and CAD processing are among the new jobs being outsourcing. By 2015, business analysis firm Forrester Research expects some 3.3 million high-tech and service jobs to be transplanted from the U.S. to other countries.

Increasing Collaboration

Because more companies are choosing to shift operations offshore, sharing data is becoming a greater priority. This has contributed to the trend toward collaboration as a management approach, and as a result, CAD vendors have been charged with the task of incorporating tools that can help different departments work together. "It's a very big challenge to build in ways to tie in other people," Robert Kross, vice president of Autodesk's Manufacturing Solutions Division, tells Design News. "Traditionally, our tools are for technical people who are full time users. Collaboration means we have to tie in other people, financial people and others who aren't doing design every day. That adds to the challenge, but it's a necessary component for today's business."

Mass Customization

Companies are now placing more emphasis on reducing inventory costs and applying just-in-time manufacturing, a strategy that is paving the way for the mass customization of products. Many firms are no longer producing huge quantities of the same product store in many warehouse which manufacturers did just five years ago. Instead they're tailoring products to meet customers' specific needs in order to become more competitive.

Popularity of 3D Modeling

More and more engineers are making the switch from 2D software to 3D modeling. In particular, solid modeling-which depicts product designs via electronic 3D solid models-is taking off. Not only do solids give engineers more design flexibility, they also offer realistic images of products and allow downstream tasks, such as analysis, to be more easily integrated. What's more, solid modeling supports a decades-long push by the CAD industry to give manufacturing customers "art-to-part" capabilities-bringing a design from the concept to the tooling stage in a completely digital world. By implementing a purely digital process, manufacturers can foster more effective communication between departments, make better products and enjoy greater profits due to a speedier time to market.

Widespread use of Neutral Formats

An increasing number of companies are exchanging information with one another using non-CAD formats, including Adobe's PDF (portable document format), Autodesk's DWF (design web format) and SolidWorks' EDRW (eDrawings). Although translating a CAD file into a non-CAD

format may seem an unnecessary step, two developments are driving this trend. Firstly, increasing globalization of manufacturing is leaving many companies susceptible to the theft of intellectual property. In response, many are refraining from releasing native CAD files and taking the precaution of employing a data format that will prevent information theft. Secondly, while digital signatures are now legally binding, many companies still doubt their effectiveness and are thus using representations that can't be edited.

In the 1970s, CAD system adopted the model which based on small machine. In the 1980's, CAD was based on the working station which lied separated, and it accelerated the progress greatly year after year. In the 1990's, CAD had a development as follows:

1. The hardware of CAD developed into a PC from the workstation. The colorful vision system and accelerate card has become the common equipment of working station and PC.
2. Various methods of model description has appeared, and gradually become standard. The research and realization of database have become more and more important. The integration system based on design has become more and more available and popular.
3. CAD technology gradually becomes intelligent, which makes the CAD system quite agile, easy-dealing, efficient and creative.

CAD can increase the speed and quality of the design, so that it can obviously effect on the economy. If we can combine the design, produce, craft processing with management to go on the integrated process, the efficient and quality of the product can make a great progress. The two stones in the way of traditional CAD system application are: the lack of intelligence and the integration in the application environment. The research of intelligence and integration is an urgent problem to be solved in the next generation of CAD system. It demands that the CAD system can provide a support (the integration of system and knowledge) to the designer, and have a description of unique and common data, that is an integration of expression model, and it should have the intelligence to find out the error and to answer the questions and give suggestions.

Development and Characteristic of ICAD

Development of ICAD

CAD has undergone several stages of development. They are illustrated as follows:

1. The traditional design technology basing on manual manufacturing and conventional CAD is the first stage of CAD. It has the main feature of product structure and performance analysis as well as computer-aided drawing. It can finish the calculation work in data calculating and picturing, but for the reasoning in knowledge model based on symbol and the symbol processing work, it is not capable. In the design, there is a lot of work needing the designer and professors to create and use the various knowledge and experience to analyze, determine and estimate, and then get the result.
2. The modern design technique basing on ICAD is the second stage. The intelligent activities are finished by the Design-Typed Expert System. This kind of system can imitate the process of the expert design, and adopt the symbols reasoning technology of single ken to solve specific issues in the single field. Because this kind of technology has limited ability, so the size and complexion are limited. This belongs to the primary process of ICAD. This kind of system combines the intelligent skills and optimization, computer-aided drawing with

finite elements, which makes the computer takes part in the programmable decision-making, performance analysis, and other conventional design process as much as possible. The efficiency increased greatly with the help of computer.

3. The advanced method basing on I^2CAD (Integrated Intelligent CAD) is the third level. I^2CAD is based on the ICAD system, and combined with various intelligent design methods. It can provide the system integration for each process, and has the only and common integration of data description. It has the intelligent features, such as finding the errors, producing the proper methods and so on. Besides the intelligent interaction interface between human and computer, it can also obtain the data automatically and provide the design methods, and can show the process and result of design intelligently. At last, the inside of the system can not only realize the network, but also compose the CAD information internet in the trade.

Characteristic of ICAD

Feature of ICAD

ICAD is an integrating research which combines AI and CAD, and it put the theory and technology into CAD, making the CAD system has certain intelligence and thinking methods of designers, and in that way, they can go deep into the automatic design, which is an important part in CAD technology. ICAD is an intelligent labor that combines discursion, colligation, association, analogy and study, and it puts the intelligent theory and methods into the CAD area. For the intelligence, there is a definition: intelligence is a system that works properly in the uncertain environment. Here the proper action means it can increase the probability of success in supporting the system to get final success. In all, we can divide the procedure into three kinds: intelligent creativity, analysis and design.

The traditional CAD system can help to finish the last two kinds of tasks: in the analysis, for example, take finite element analysis technology as the representative, the skills of analysis including all kinds of professional system including calculating, imitating, analyzing and so on; in the show of product, including the early CAD technology, planar draw, three-dimensional skills and the extra benefits it brings, such as the examination, assembling and so on. The key of design is the intelligent creativity, for which the traditional CAD skills cannot provide effective assistance and support.

The ICAD research is to establish an intelligent system or environment that can help the designers to finish the design including intelligent creativity, so that it can improve the degree of automatism. In this way, ICAD is an extension of traditional one, and it is a true and full sense of the computer-aided "design". It is need to emphasize that the ICAD is not aiming on "design automatically". On the one hand, design is a really complex activity. There is not a clear and whole understanding about the design and how it can use computer to imitate and take the place of human beings. Only focusing on "design automatically" will definitely lead to the narrow usage about the automatic design system and being unsatisfied to the various real needs. On the other hand, computer system and human beings can form an interaction in the design. Computer cannot deal with some problems which are simple for human. But the designer can take part in it and make it easier and can examine the mistakes in the procedure in CAD system, so that it can avoid misleading the design. In all, the ICAD insists that computer is in the assistant place in design, human is also needed to take part in the design and play the main role.

All in all, depending on the objective and realistic attitude, the goal of ICAD system should be as follows: establishing systems or environments to assistant designers to accomplish design tasks including intelligence creation, analysis and calculation, product expression, and so on, improv-

ing automatization level of design and releasing human being designers.

Features of I^{2}CAD System

In 21 century, the information age and internet age have come into our life. Intelligent design has made a great progress. As the core of design skill in the new century, I^{2}CAD has its new features.

1. **Integration**
   - **The integration of facing the entire product life cycle:** I^{2}CAD system combines the design with producing, and it also pays attention to the materials including the stockpile management, producing program, finance management, sale and so on. It can combine the design, analysis, machining, assembling, testing and management, and also it can reduce the cost and shorten the time spent on design and produce through the communication and exchanging of computer and the well cooperation of each section.
   - **The integration of intelligent methods:** the problem of design is about single input and multi-output, that is, the customers demand the specific product, while the system can achieve it via many different methods and after comparison, they will get the best one. I^{2}CAD system may integrate the various intelligent design methods, such as intelligent optimized design, object-oriented design, intelligence-oriented design, concurrent design, collaborative design, information flow design and virtual design
   - **The integration of design knowledge:** the procedure of design mentions various application of knowledge, including empirical, common-sense and structural knowledge. The intelligent design should be based on many kinds of design knowledge. I^{2}CAD system can achieve the application, integration and management through the establishment, management and combination of knowledge database.
2. **Intelligence**
   - **The redesign and self-study mechanism of I^{2}CAD:** The main feature of intelligent system is to redesign and self-improve automatically on the condition that human intervention is minimized. That is to say, if the result of the design cannot satisfy the demand, the system can redesign and use the error message, the stored knowledge in the database and the dynamic response message from the customers to design the feedback, finish the redesign task and also adopt the concluding, reasoning and analogy to get new knowledge and enlarge the database, go on self-study and self-perfect. So that the computer may probably take the place of the designer in maximum.
   - **The intelligent interaction interface of human and computer:** The excellent interaction is necessary to ICAD. I^{2}CAD system should achieve the recognizing, choosing, understanding and consequence, so that the interaction will be simpler and friendlier.
   - **The visualization of system:** As the visualization technology, virtual reality technology and the intelligent multi-media technology develop, I^{2}CAD will show the process and the result of the design intelligently. For example, combining the multi-media technology, imitating the facial features to accept and deal with the information through words, figures, pictures and sounds. CAD virtual environment can make the designer stay

in the space designed by himself, and have a close feel to the product and engineering design and layout, and can exert the designer's intelligence to make the design perfect.

3. **Automatization**
   - **The automatic generation plan.** In I^{2}CAD, from the surface to feature design, from the conception design to the procedure design, the designer demands the system to take the place of professor and produce the programs which can satisfy various demands automatically. But the theory of design is not mature till now and there are a lot of problem waiting for us to solve. The automatization design is hard to realize. The more realistic method is to produce local automatic programs or interaction between human and computer.
   - **Obtain the data automatically.** The blueprint is the main media in the project, it records amounts of information. Obtaining the data automatically means the blueprint being input automatically and intelligent recognized, that is, by scanning the lattice images and the intelligent interface in the CAD system, based on the lattice images and the vector pictures changing automatically, it can achieve the recognization and understanding of the blueprint, and transfer it into the data form that can be compatible in CAD system.
   - **The automatic establishment of 3-D shape.** The automatic establishment of 3-D shape is to produce the 3-D objects automatically through the 2-D geometry and topology information in 3-D views. The aims include two sides: on the one hand, it provides I^{2}CAD a new method of 3-D model. On the other hand, it offers an efficient and dependable way of 3-D model establishment in the process of transferring from design program to the result.
4. **Openness:** ICAD system is one that combines the cooperation between human beings and computer. The opening of I^{2}CAD system means the system should be operable, that is, the system should provide the designer a flexible platform to do free design for themselves, and the system should be redesignable.
5. **Network:** Engineering technology is in close connection with computer and communication technologies in information age, which makes it necessary to realize the automatic design connected to the network. The network of I^{2}CAD has the following two methods:
   - **The local area network (LAN).** I^{2}CAD constitute a LAN by itself. All of the public information in the design, such as picture, words and coding, should be stored in the public database. Each of the working station shares the data inside through the internet and design. Each working station also exchanges the result through the internet.
   - **By internet.** Every I^{2}CAD can make up CAD information net through the internet. The designers and decision makers can get many kinds of information through the internet, and associate many I^{2}CAD systems to solve a complex problem of design together and may form a system of "Internet Design Institute".
6. **Sustainable development:** The aim of sustainable development strategy in manufacturing field in the 21st century is to save the resource and prevent the environment pollution. So the research and design of the

product should not only consider the theory of task and function of structure, but also should mind the saving resources and preventing pollution. Just for this, many kinds of design is developing, such as design for environment (DFE), design for disassembly (DFD), design for recycle (DFR) and so on. Those theories and methods are the new way of development, and can play an important part in the design area.

RESEARCH AND APPLICATION OF ICAD

Current Research

Presently, famous research institutes in the world engaging in CAD study primarily concludes.

Carnegie Mellon University USA. During the mid 1970s, research had being carried out on living space synthesis method under the lead of famous artificial intelligence scholar H.A.Simon and famous CAD scholar C.Eastman. The above-mentioned school has set up “the engineering design research center of cross-department ” in 1986, and studies in intelligent behavior of human being design activity was carried out by the scholars of such field as in computer, mechanical engineering, civil engineering and architectural engineering from different levels of proposed design ideas, design methods, the change of proposal and the evaluation of design respectively. So far, they have already developed a lot of automated synthetically expert systems or intelligence design environment schemes used for structure scheme, interior decoration scheme, and electromechanical device scheme. University of Sydney Australia. Designing activity was transformed into artificial intelligence technology study by the computational institute led by J.S.Gero professor from late 1970’s. At present, they leads the world in design theory, case reasoning, cognition and learning process in designing, designing knowledge base, designing the expression of grammar, innovation design model, design process, etc. Artificial potassium system of the University of Edinburgh. United Kingdom also achieves remarkably in the research of ICAD whose job is to build the ICAD environment to support person making designing and it was thought more realistic compared with the establishment of a person doing the ICAD system designing. Their job started from 1984 and came to the end in 1990 as a part of Design to Product in CAD/CAM plan which cost ￡9200,000,200 man/month. And AI department in the University of Edinburgh has brought forward an exploration-based design model concentrated on the arrangement of designing knowledge. University of Tokyo. Japan. Professor H.Yoshikawa has brought forward general design theory (GDT) in 1977 which has established the important basis of intelligence CAD studies. Intensive study in the application of artificial intelligence technology in unfinished products designing scheme has been carried out by his cooperator Professor T.Tomiyama who has set up a Tomiyama laboratory mainly engaged in the research of intelligence CAD, function design and knowledge frame of product design. National standard and technology institute manufacturing engineering laboratory. USA. The laboratory nonscheduled provides jobs for the world’s famous scholars in intelligence design field. Research subjects currently engaged are knowledge expressing model in product designing, next-generation studies of knowledge designing base in distributing design environment and so on.

Application Examples

I^3CAD

A number of prototype systems have been developed for industrial companies. The I^3CAD system was developed for the New Japan Steel Corporation. The feature-based product modeling system was modified as an application for

Gienow Building Products Ltd. for designing building products.

In this research, design process knowledge is represented at two different levels, action level and object level, corresponding to the meta-knowledge to model design behaviors and the special knowledge to model the processes in designing particular objects. A design knowledge base and database modeling language-Integrated Data Description Language (IDDL) was developed at the University of Tokyo to model both design processes and design objects. This language combines logic programming functions and object oriented programming functions into an integrated environment. Using this language, an ICAD system - I^3CAD was developed at the University of Tokyo. Contradictions of knowledge base and database are resolved using circumscription and Assumption-based Truth Maintenance System (ATMS) in this system. Many advanced knowledge modeling techniques, including Qualitative Process Theory (QPT), modal logic, default reasoning, etc., have also been introduced in the I^3CAD system. The knowledge base and database representation scheme of IDDL serves as the basis in the feature-based integrated concurrent design system. In the integrated concurrent design system, a new feature modeling language was introduced. In addition to the qualitative and quantitative data/relations that were introduced in IDDL, the composing element features, data dependency relations, constraints, 2D and 3D feature geometric descriptions have also been introduced for representing product life-cycle models and their relations. Optimization functions were added to the integrated concurrent design system to identify the optimal design considering relevant life-cycle aspects.

An ICAD System For Multi-material Layered Manufacturing

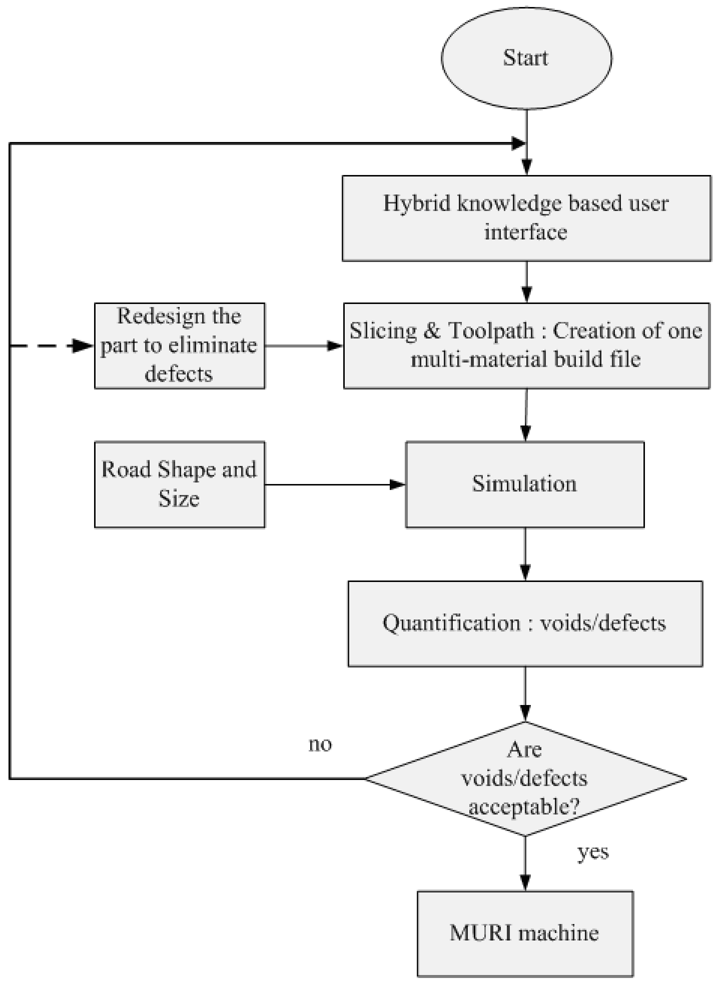
Layered Manufacturing (LM) refers to the process of fabricating three-dimensional objects from CAD-generated solid models layer by layer. Since the first commercial inception of Stereo lithography technology in 1988, the LM technologies have been developed rapidly. The key benefits of LM are "It's faster, cost less, and quality is improved." And this claim can be supported by success stories giving specific comparative examples of time and cost savings relative to the more traditional ways of completing the same steps in the product development and commercialization cycle.

At Rutgers University, under an Office of Navy Research (ONR) funded MURI (Multi-disciplinary University Research Initiative) program, An Intelligent Layered Manufacturing System for fabrication of multiphase electromechanical parts has been developed. This program will requires development of design and construction of an intelligent multi-material LM system, processing science for multi-material LM, CAD based design and manufacture of multi-phase electromechanical components. The intelligent multi-material LM system should include a high quality multi-material layered manufacturing machine (hardware), and a CAD based intelligent multi-material LM software system. In the recent past, the MURI hardware as well as corresponding software has been developed and several multi-material parts have been successfully fabricated. This research concentrates on the intelligent multi-material CAD system.

The ICAD system under development includes: solid model design, generation of .stl file, formation of multi-material tool path, virtual simulation and defect quantification of multi-material LM. As shown in Figure 1, this CAD process is a design-analyze-redesign process.

The goal of the CAD system is to virtually fabricate high quality multi-material part and have the multi-material toolpath ready for MURI machine. The solid model design and .stl file format generation can be done using commercial CAD software like I-DEAS, ProE and AutoCAD. The in-house software being developed here are multi-material

Figure 1. ICAD system for multi-material LM system



toolpath generation, virtual simulation and defect quantification of multi-material LM.

The multi-material CAD system in an intelligent Layered Manufacturing System is to generate toolpath data file ready for the MURI machine and to virtually fabricate defect free multi-material parts. In the present program, quality of a part is considered to be very important. To have a full control over the building toolpath, which has important influence over the part's quality, an in house intelligent toolpath generation software is needed. The objective of this study is to develop a virtually void free toolpath file for the multi-material LM process.

The Multi-material CAD system (Figure 1) includes multi-material build file generation, virtual simulation and quantification of the defects in virtual part. The virtual simulation uses the in-house multi-material build file as the input data and virtually fabricates the multi-material part. Surface and internal defects can be quantified from the images captured from the outside view or sectional view of the simulation. If the part quality is not acceptable, modifying the toolpath parameters to regenerate the build file or redesigning of the part is needed. The iteration runs until the virtual part quality is acceptable. Finally

the multi-material build file here is ready for the hardware - MURI machine.

An ICAD System for Cold-Forging Parts in Automobile Industry

Cold-forging technology has developed rapidly, so it is very important to advance the cold-forging process and die design, which has been heavily dependent on the experience of engineers. The great development of AI technology made it possible to build up intelligent systems to resolve problems requiring qualitative and ambiguous human experience.

In recent years, several CAD systems were developed for solving various technological tasks. In these – often called "conventional" CAD system, some numerical calculation and graphical capabilities can be performed easily by computer, but the process planning – often requiring expertise - is still realized by interactive inference of the user. The great development of AI technology made it possible to utilize the successful experience of experts to build up intelligent systems to resolve problems requiring qualitative and ambiguous human experience.

The knowledge in cold forging process and die design, according to the industrial technology, can be classified into several groups: forming, materials, machines, etc. However, in order to establish an intelligent system, it's necessary to make a classification according to knowledge engineering. From this point of view, the knowledge can be classified into three main groups: facts, procedures, and heuristics.

Facts means the basic part of knowledge, which precisely inform you the quality or quantity representation of some important features, such as material properties, product geometry, etc.

Procedures include rules and numerical methods, which can change geometrical and technological data information into other data information, namely other facts. The group of knowledge is the key issue of process and design, without which the designing work in computer is impossible to perform.

According to the engineering question, some complicated nonlinear relations between input and output data are wanted. So, a newly-developed AI technology - artificial neural network (ANN) is included in procedures to simulate the prediction of cold-forming elastic deformation, the die life, and some other key factors in die design. Meanwhile, some conventional numerical analyzing methods are also represented as procedures.

The process planning of multi-stage forming processes is a very crucial problem in cold-forging process. The same and similar parts can be produced using a large number of different feasible processes. So it's somewhat difficult to select the most reasonable process. The character of process planning made it necessary to use heuristics to represent such kind of knowledge.

These three groups of knowledge are stored in the knowledge base in a certain structure. The input data for the system includes the final product geometry and material type. Material type can be represented by character l string in computer directly, Coding, however, is very important for the product geometry's representation.

To perform process and die design, the prediction of elastic deformation and the die life are crucial and difficult. For cold-forging, the aim of which is to produce near-net shape or net shape product, the elastic deformation is very important for keeping the precision of final products and the die life is very crucial for evaluating cost of a forming process. Because there are many factors related to them, and the relations between the factors and them are complex and nonlinear, it is necessary to introduce ANN method to resolve the difficulty. At the same time, some ambiguous knowledge (such as the complexity of product) needs fuzzy mathematics model to represent. So, fizzy neural networks are applied to establish prediction model of elastic deformation and die life.

SolidWorks

A robust suite of core productivity tools, SolidWorks Office Professional includes SolidWorks 3D mechanical design software, a full range of design communication and CAD productivity tools and PDMWorks, an easy-to-set-up-and-use product data management (PDM) solution that is uniquely adapted to managing SolidWorks product data for the individual or workgroup.

3DQuickPress is a SolidWorks® add-on for progressive die design. With a 3D solid model of the part, a 3D strip can be created quickly and easily, and also be communicated clearly throughout the work group using the free SolidWorks eDrawings™ viewer before further detailing of the die without time wasted.

Powerful sheet metal feature recognition technology can handle native SolidWorks® sheet metal parts or imported data, and provides a knowledge base for spring back and bending allowance. Strip layout is the core of a progressive die design. The powerful Strip Layout Manager provides users with intuitive tools to quickly complete the strip simply by drag & drop command with instant graphical feedback.

3DQuickForm is the latest powerful SolidWorks® add-in application for inverse forming simulation. High speed, accuracy and full associatively with SolidWorks® data are taken into consideration to make 3DQuickForm™ a production design tool for die designers. Users can either import different types of CAD data or build the die geometries in SolidWorks® environment for the forming studies. It helps OEM manufacturers, material suppliers and die manufacturers to deliver precision tooling in the shortest time and with minimum physical try-outs.

The built-in mesher is automatic and allows user adjustment to meet needs in different situations. Tool designers may fully concentrate on the die design processes and virtual try-outs. Simulations are run in project driven manner and support multiple projects, thus streamlining design feasibility study, process simulation and optimization. Simulation results are displayed to show initial blank shape, different material thickness, and stress/strain distribution. Customizable material database allows designers to test their design with different materials. Fully integrated with SolidWorks, 3DQuickForm™ is easy to use and rich in functions.

The powerful version of 3DQuickForm 3DQuickForm™ Advanced, provides also stamping and flanging simulation, initial flat or initial curved bland, blank holder definition, various boundary conditions like fixed edges/faces, drawbeads, and symmetric plane, etc. Multiple-steps is supported as well.

Thinking Software Vellum 3.0

CAD software is looking forward to be intellectualized. A company in Shanghai, which is the general agent of Ashlar, has fetch the Vellum in China, and it has been showed on the exposition. The software is called “thinking software”, it can know the author’s purpose, which represents its intelligence. For example, if you draw a beeline, at the end of the beeline, the software can clue to you the existed relationship of the graphics, automatically catching the required position, telling the manipulator to choose the midpoint, the end-point, the tangent of the arc, the middle of circularity and so on at any moment.

Vellum’s 2D, 3D function can mutual cooperate, for those mechanism, they can draw a 3D or a actual graphics of a accessory in a few minutes, the only thing they will do is drawing a floor plan on the screen, giving out the thickness, when that had been done, a 3D graphics is finished. The company has alleged that many software companies were trying to research software, but all were failed. Vellum is successful. Surface is constructed n automatically and the shape of aperture, rough, curve is understood easily.

Unlike many other 3D CAD programs, Vellum allows the designer to work in the z axis as

he creates a model. Rather than drawing lines in front, top, and side views as in other programs, the user simply draws the design in Vellum's Trimetric view, which is similar to sketching it on paper. This program differs from other mid-range CAD systems in that it includes a strong surface modeling capability.

Vellum 3D is a sophisticated 3D CAD program that bridges the gap between 2D and 3D. With a single menu pick, Vellum 3D transforms 2D and 3D wireframe data into surface data which can be rendered for visualization and detail drawings. Another menu pick and volume information can be sent to Vellum Solids, Finite Element Analysis or Computer Aided Manufacturing packages, etc. Vellum 3D features robust 3D wireframe modeling, associative views, associative dimensioning, parametric modeling and detail views dynamically linked to main part drawing. Always accessible, smooth handling and flexible, Vellum's advanced tools add power and leverage without compromising the program's easy drawing features that speed the design to completion.

The function of model and modification of the video 3D object can compare to the workstation software that showed on the 30th geologic international conference. Though this software can be run on Pentium computer which can be achieved easily, it is not difficult to configure such hardwares.

TECHNIQUE AND RESEARCH METHODS OF ICAD

Design-Typed Expert System

Expert systems bring a lot of advantages to designers as being one of the application areas of artificial intelligence. These systems can be used widely where expert knowledge is needed. Problems are solved with the software developed by expert systems without an expert or with a little aid of expert knowledge. Experts possess a certain accumulative amount of knowledge and experience in their area. If software developed by expert systems could use such experience and knowledge then these programs could also have high performance in related areas. It is planned to rice productivity using the expert systems at the stages of design manufacturing, marketing and evaluating customer requests (Kayır Y.& Gülesin M., 1996).

At the sectors of machine and manufacturing, expert systems are widely used in decision making based on experience and knowledge. Expert systems could also be used successfully to automate some production stages that require reasoning. Since expert systems automate reasoning procedures, capture expertise knowledge and evaluates the domain knowledge, they are more efficient than conventional programming techniques (Paszek R. & Knosala,1997).

Available CAD systems are not sufficient for all of the necessities of the design. As a result of this reason many designers believe that CAD systems should have more strength and be able to do more. Due to developments of artificial intelligence (AI), the expectancy from CAD technologies has risen. Interface mechanism of artificial intelligence technologies knowledge bases, research methods, etc., make computers more intelligent in solving problems. A lot of researchers have started to use artificial intelligence technologies in developing of CAD systems. Many experts developing and using expert systems to solve engineering problems. CAD systems are used widely nowadays. If a problem cannot be solved with algorithmic techniques and system is meant to be working like an expert then artificial intelligence is used.

The developments in AI techniques affected CAD systems. By adding AI features to CAD systems ICAD (Intelligent Computer Aided Design) systems have been developed. In spite of all developments in this area ICAD systems have not yet reached to the expected level. Knowledge based engineering systems are developed in order to determine complex model and structures with or

without graphics. When ICAD systems are being developed, information is gathered from engineers and then files is generated for customer needs based on this information. Thus organized and detailed data are obtained. This allows developing more suitable design with the knowledge.

There are two main methods in ICAD systems. In the first approach main strong design knowledge is used. ICAD should be able to overcome complex design knowledge. The system differs with natural design systems and form a skeleton concerning general structure. After that procedure analyzing of parts is done. In this system ICAD gives answers to the questions about which elements will be used in the skeleton. In the second approach by narrowing definition set of the system, the first part of the system is built. In this way the whole of the system is developed part by part. Using analyzed results this approach clarify, which design knowledge and expert system or knowledge based design system should be used.

Using artificial intelligence techniques, several CAD systems such as TIPPS, DICAD, IIICAD have been developed. TIPPS, uses artificial intelligence and concept of decision tree. The system uses a special language named process knowledge information (PKI) and B-rep CAD models as part data. Procedure information system is used for milling and drilling operations. To determine manufacturing operations, operation parameters and time span for manufacturing. The user only should select the surfaces on which operations will be applied.

Ouyang et all., applied artificial intelligence techniques in model design for tools (Ouyang M.A., Li C.G., Zhong Y.F., Yu J., Zhou J., 1996). The researchers explained the methodology of conceptual modular design and investigated the probability of combining conventional CAD systems with this new approach. Building design model and system structure, the techniques were used with expert systems, condition based recognition and neural networks.

Dialog Oriented Intelligent CAD (DICAD) has been developed at Karlsruhe University by Grabowsky and his colleagues. DICAD is concerned with the integration of geometric information with product model, functions and structure.

I^3CAD has been developed in Amsterdam at the center of computer and mathematics. In the system relation between object and operation information is defined. Meta model is used in I^3CAD. Meta model, is a central design model which can be presented in application model. I^3CAD uses Artifact Design Description Language (ADDL). Using this language design information and data information can be defined.

Considering studies made on ICAD it is clearly seen that expert systems have very important roles and it is also obvious that it will make very important supplements to future studies of ICAD.

Building an expert system is known as knowledge engineering and its practitioners are called knowledge engineers. The knowledge engineer must make sure that the computer has all the knowledge needed to solve a problem. The knowledge engineer must choose one or more forms in which to represent the required knowledge as symbol patterns in the memory of the computer, that is, he (or she) must choose a knowledge representation. He must also ensure that the computer can use the knowledge efficiently by selecting from a handful of reasoning methods. The practice of knowledge engineering is described later. We first describe the components of expert systems.

Every expert system consists of two principal parts: the knowledge base and the reasoning, or inference, engine.

The knowledge base of expert systems contains both factual and heuristic knowledge. Factual knowledge is that knowledge of the task domain that is widely shared, typically found in textbooks or journals, and commonly agreed upon by those knowledgeable in the particular field.

Heuristic knowledge is the less rigorous, more experiential, more judgmental knowledge of performance. In contrast to factual knowledge, heuristic knowledge is rarely discussed, and is largely individualistic. It is the knowledge of good practice, good judgment, and plausible reasoning in the field. It is the knowledge that underlies the "art of good guessing."

Knowledge representation formalizes and organizes the knowledge. One widely used representation is the production rule, or simply rule. A rule consists of an IF part and a THEN part (also called a condition and an action). The IF part lists a set of conditions in some logical combination. The piece of knowledge represented by the production rule is relevant to the line of reasoning being developed if the IF part of the rule is satisfied; consequently, the THEN part can be concluded, or its problem-solving action taken. Expert systems whose knowledge is represented in rule form are called rule-based systems.

Another widely used representation, called the unit (also known as frame, schema, or list structure) is based upon a more passive view of knowledge. The unit is an assemblage of associated symbolic knowledge about an entity to be represented. Typically, a unit consists of a list of properties of the entity and associated values for those properties.

Since every task domain consists of many entities that stand in various relations, the properties can also be used to specify relations, and the values of these properties are the names of other units that are linked according to the relations. One unit can also represent knowledge that is a "special case" of another unit, or some units can be "parts of" another unit.

The problem-solving model, or paradigm, organizes and controls the steps taken to solve the problem. One common but powerful paradigm involves chaining of IF-THEN rules to form a line of reasoning. If the chaining starts from a set of conditions and moves toward some conclusion, the method is called forward chaining. If the conclusion is known (for example, a goal to be achieved) but the path to that conclusion is not known, then reasoning backwards is called for, and the method is backward chaining. These problem-solving methods are built into program modules called inference engines or inference procedures that manipulate and use knowledge in the knowledge base to form a line of reasoning.

The knowledge base an expert uses is what he learned at school, from colleagues, and from years of experience. Presumably the more experience he has, the larger his store of knowledge. Knowledge allows him to interpret the information in his databases to advantage in diagnosis, design, and analysis.

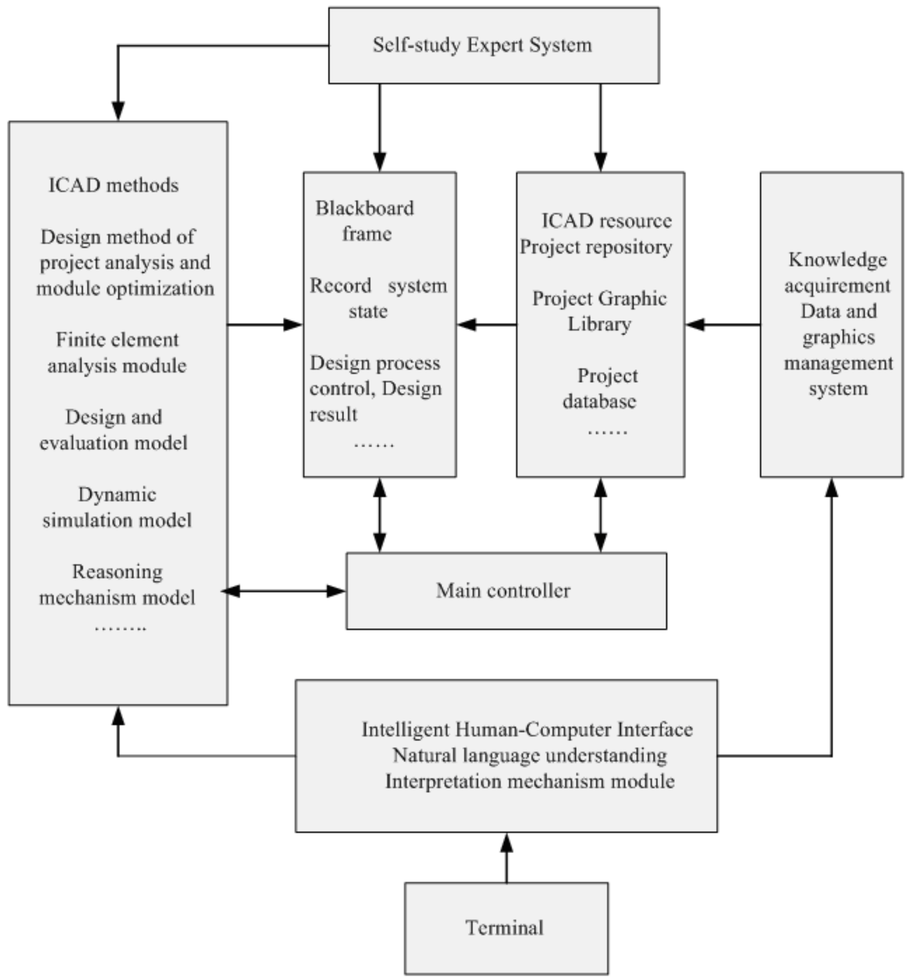
Though an expert system consists primarily of a knowledge base and an inference engine, a couple of other features are worth mentioning: reasoning with uncertainty, and explanation of the line of reasoning. Figure 2 shows the architecture of Design-Typed Expert System.

Knowledge is almost always incomplete and uncertain. To deal with uncertain knowledge, a rule may have associated with it a confidence factor or a weight. The set of methods for using uncertain knowledge in combination with uncertain data in the reasoning process is called reasoning with uncertainty. An important subclass of methods for reasoning with uncertainty is called "fuzzy logic," and the systems that use them are known as "fuzzy systems."

Because an expert system uses uncertain or heuristic knowledge (as we humans do) its credibility is often in question (as is the case with humans). When an answer to a problem is questionable, we tend to want to know the rationale. If the rationale seems plausible, we tend to believe the answer. So it is with expert systems. Most expert systems have the ability to answer questions of the form: "Why is the answer X?" Explanations can be generated by tracing the line of reasoning used by the inference engine.

The most important ingredient in expert system is knowledge. The power of expert systems

Figure 2. Architecture of design-typed expert system



resides in the specific, high-quality knowledge they contain about task domains. AI researchers will continue to explore and add to the current repertoire of knowledge representation and reasoning methods. But in knowledge resides the power (Lin P., 2006). Because of the importance of knowledge in expert systems and because the current knowledge acquisition method is slow and tedious, much of the future of expert systems depends on breaking the knowledge acquisition bottleneck and in codifying and representing a large knowledge infrastructure.

AI comes into CAD and becomes an important form of ICAD ie. designing expert system. Normal method is to use in the exsit expert system, researching a system aim at some certain design. Expert system is a software that provide knowledge express method, knowledge management system and study ability. When the user properly express the knowledge, and then a system that can solve the problem is established. Some CAD researchers have create special CAD expert system, but now the expert system always localize in the part of logic thinking, such as a request of a problem, a design of a process. It is for the reason that the expert system is based on logic thinking, for the core problem of design, that is shape integration, still lack of effective approach. so design-typed

expert system still need to improve, so that it can be applied in the core process of CAD.

There are many non-structural or semi-structured problems in the engineering design. There is not proper arithmetic and must use experience, logic reasoning and judgement to solve the problem. Due to the discrete and non-systemic of knowledge, the knowledge, judgement and decision of experts are different, sometimes their decision is not perfect. Thus, the design-typed expert system based on knowledge engineering technology is valued. On the other hand, the flow of the product design usually has the character of big decision-making room, multi-ply targets and multi-ply restrict. Any product design task can be analyzed from different aspects, and have a decomposition of the overall objectives, functions, tasks and knowledge layer by layer, then it can be expressed by tree structure. This requests the expert system's repository to use layered structure, the bottom is the smallest cell of the knowledge module, and the solution is divide into some sub-missions, according to the mission,. The expert system has great potential in engineering design, it is the tool and method to deal with the engineering design. It can make interpretation, evaluation, detection, so that users can make optimal decision. In addition, it is also provide good solutions to random and uncertain engineering problems.

The most distinct character of the expert system is the software has solution repository. It has the reasoning decision-making mechanism which can select knowledge and coordinate project databases and graphics library resources to complete the design tasks together. So except the CAD components, the design-typed expert system should also have intelligent module such as knowledge base and reasoning mechanism.

Agent-Oriented ICAD System

The CAD system running under the stand-alone environment does not have the communication function and the management of the coordinated activity mechanism. Meanwhile, the CAD system is on its way from single-user-orientation to multi-user-orientation. Thus the corresponding CAD software should be applicable to this kind of group collaborative design. As the popularization of and advancement in CAD technology goes on, CAD users will have more and more design data which is difficult to manage. At the same time it is necessary to offer enough network information for the follow-up link such as the craft management, etc. CAD system should be able to run on Internet and take full advantages of the Internet resources.

In order to meet these requirements, it is necessary to have efficient collaborative design environments. These environments should not only automate individual tasks, in the manner of traditional computer aided engineering tools, but also enable individual members to share information, collaborate and coordinate their activities within the context of a design project.

As an emergent approach to developing distributed systems, agent technology has been employed to develop collaborative product design systems by a number of researchers (Coad P. & Yourdon E., 1998). Agent comes from the field of distributed artificial intelligence (DAI). Some representative definitions of agent are as follows: Wooldridge and Jennings in Britain hold that the term agent is generally used to denote a hardware-based or more often a software-based computer system that enjoys the following properties: autonomy, social ability, reactivity, pro-activeness .

The agent view provides a level of abstraction at which we construct computational systems that inter-operate globally across networks linking people, organizations and machines on a virtual platform. The term 'agent' means an 'entity' that is autonomous, behaves intelligently, moves around over the web and communicates and cooperates with other agents. A multi-agent system consists of autonomous interacting intelligent agents. In this approach, the distributed computational agents are

knowledgeable in their local domain and share the responsibility of achieving multi-objective system goals through concurrent negotiation.

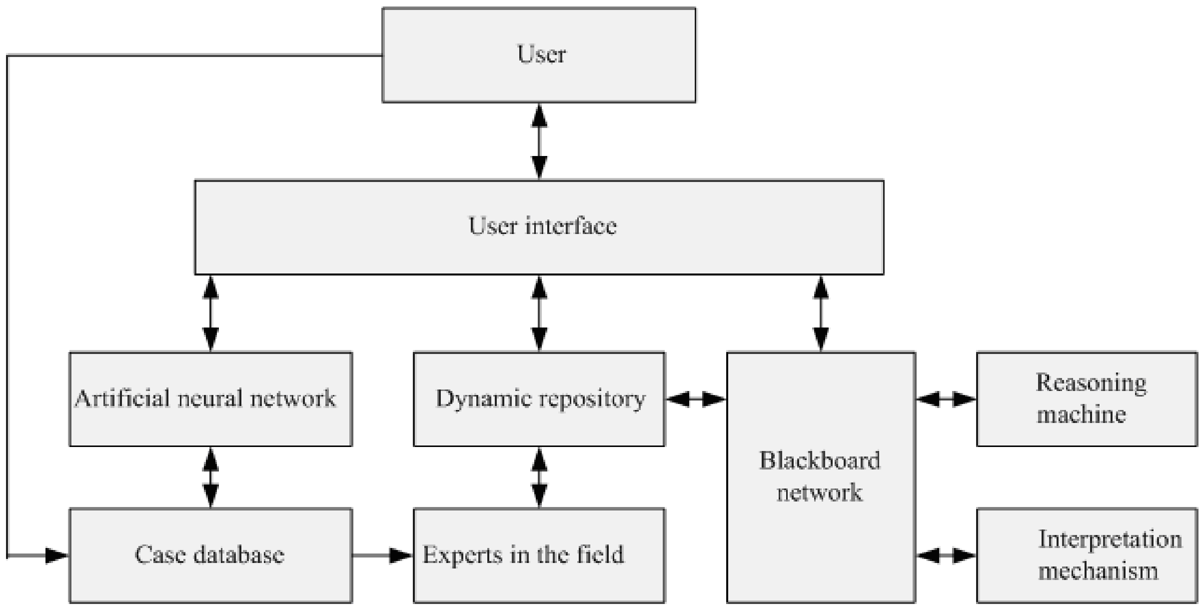
In the past years, a number of researchers applied agent technology to concurrent engineering, enterprise integration, distributed process planning and production scheduling, etc. PACT was a distributed concurrent engineering system based on interacting engineering tools that were wrapped up as agents. The SHARE project was looking at an open, heterogeneous, network-oriented environments to help engineers and designers collaborate in mechanical domains (G. Toye M.R., Cutkosky L., Leifer J., Tenenbaum J. Glicksman., 1993). FIRST-LINK, NEXT-LINK, and Process-Link were continuous projects to propose, develop and test agent-based systems for integration, coordination of distributed CAD tools. Balasubramanian et al. proposed a multi-agent approach to concurrent design, manufacturability analysis, process planning, routing and scheduling. A heterogeneous multi-agent concurrent engineering system consisting of multiple feature-based design sub-system, multiple simulated shop-floor resource groups, a supervisory control interface and the coordination mechanisms for multi-agent cooperation, has been developed. Distributed intelligent design environment (DIDE) was a distributed architecture for integrating multidisciplinary engineering tools in an open environment, organized as a population of asynchronous cognitive agents (W. Shen, J.P. Barthe`s.,1997). Based on DIDE, Meta-Morph I and Meta-Morph II were continually developed for agent-based distributed intelligent design and manufacturing. Sun et al. proposed a distributed multi-agent environment for product design and manufacturing planning. The main objective was to develop a distributed concurrent engineering system to allow geographically dispersed entities to work cooperatively towards overall system goals.

Agent is a higher level abstraction in software engineering. Agent based approaches alleviate the high coupling of distributed objects by using a message passing communication paradigm. In a multi-agent system, an agent is characterized by its autonomous and cooperative capabilities, such as independent reasoning and local decision making, handling incoming messages and managing multiple tasks at the same time. However, the agent technology seems to be too complicated for practical applications without the provision of essential components that form the baseline for agent system development process. Moreover, for many reasons, agents developed by different organizations are difficult to interoperate because of the lack of a common messaging and ontology interface.

FIPA, founded in 1996, is an international standardization organization promoting the development and specification of agent technologies. The main goal for FIPA specifications is to specify how different kinds of agent platforms can interoperate. In the context of FIPA, an agent is an encapsulated entity with its own states, behavior, thread of control, and an ability to interact and communicate with other entities-including people, other agents, and legacy systems.

Recently, it is popular to use agent to construct design system, the usual way to construct system is to use many independent modules. The small modules are easily made and maintenance, but the difficulty is how to communicate among the modules and how to manage and organize lots of small modules. Agent based on blackboard structure can solve such problems, in this style of agent communication model, each agent tries its best to solve the sub-problem, in order to reduce communication traffic and improve system efficiency. When some agent can not complete its target and needs help, it sends a message to the blackboard agent and announces a protocol request, other agent who reads the message arrange an individual planning to the initial agent. every collaborative agent can make a cooperative planning and decide its action, actions of other agents and actions completed together with other agents. At the same time, each agent should con-

Figure 3. Structure of agent-oriented ICAD system



sider three factors: eliminating obstacle for other agents, achieving the work load balance of agent, and getting higher efficiency at a lower price.

Figure 3 presents the structure of agent-oriented ICAD system, the characteristic of the structure is:

1. For the part of ICAD based on the ANN, first use examples to train neural network to get weight matrix of ANN and use this part to project design. In this process, we can use new problems as study samples to amend the weight matrix of ANN. For the system stabilization, the weight matrix should be amended when the system has worked for a period of time for accumulating a certain number of study samples.
2. For the part of ICAD based on agent, when the system comes to hand new design task, effective agents can be organized to complete the task. There are two approaches:one is to create new agent, when the task is completed, the corresponding agent can be cancelled, the other is to use leisure agent to construct the execute agent aggregate. Besides, according to the task resources and external information which change continuously and periodically, as well as the past behavior of the system, to change style and number of agent in organization system for organization reconstruction.

Although the ICAD system needs to be improved, the intelligent level (or automatization level) is not considered the higher the better, because the higher intelligent the system is, the more cost it will be paid, and this growth is almost exponential relations. In fact, to some extent, it is complementary among human, expert system and ANN. Something is easy for human, but is hard for machine, even pay a lot of manpower and material resources, the effect is not good. Therefore, to establish an intelligent system of "man-machine integration", taking good use of human's intelligent advantage, making man co-operate with machine equally, paying reasonable cost to achieve higher intelligence, which will be

a guiding principle in the research of CAD system based on agent for a long time.

Research Methods of ICAD

Intelligent Design Method

The result of design can have an influence on the whole process of design, and it has an important effect on reducing the cost, developing the quality and shortening the time of design. It is the key to realize the innovation of the product. The design methods facing the function of product are the basement for the intelligent design. At the same time, the intelligentization of design also concludes the integration of fuzzy intelligent optimization, artificial neural network and evolution optimization, etc.

Collaborative Design Method

Collaborative design method is the key to assure the integration and share of many kinds of knowledge, data, information, discursion and decision. The software environment of multi-expert system for collaborative problem solving is a basic means to achieve a task or goal with coordination and cooperation of multi-expert system.

Concurrent Design Method

Concurrent design method is the important way for the intelligent design system to realize the dream of facing manufacturing design and the whole life cycle of product. That is, at the process of product design, the influence of craft programming, manufacture, assembling, testing and maintenance in product life cycle should be considered. Each tache integrates concurrently, which will shorten the time of design, reduce the cost of product and improve the product quality.

Case-Based Reasoning (CBR) Method

CBR is a new method of reasoning and self-study (Nathan D. M., 1998). The core spirit is to use the case of successful experience to solve the new problem. The basic steps of CBR are: raising question, finding out the similar case, modifying the case to satisfy the reality, and storing the new satisfactory case into the database. The most important support of CBR is the case database, and the key is the high efficient extraction of the case.

CBR is an intelligent skill to get the answer of the current question through exploring the similar answer in the database. In CBR, the condition description and the answer will be presented by a case, but the case itself can be regarded as a model and will be stored in the database as a certain form. When a new problem appears, the system will depend on the index, and recall the similar case from the database, and then modify it to make it fit for the demands of the current problem.

Object Oriented Design Method

This method is doing research on the design product which is regarded as parametric object. First of all, establish an integral model, the discretional structure can make up an assembling model, and each structure should belong to an object which has a different parameter. When transferred, it only need to change the parameter, that is, facing the different objects. This object oriented design is good for the establishment of an open and agile ICAD system.

Method Based on Integrated Reasoning

The idea of reasoning being used in the design, the formation of the program can be treated as the process of reasoning. When the data and knowledge are input into the computer, the computer will get the proper design planing through reasoning. Besides, the design of human beings

is another kind of high integrated intelligent activity, and the designer can produce new ideas through integrating various conditions, and also can use the old experience or imagination to get the result. This kind of method is called integrated reasoning method.

Prototype-Based Design (PBD) Method

The human professor can often follow their former experience to conclude problems of design into some typical forms of structure, and also when they meet new problems of design, they will adopt the typical forms and choose one of them as a solution, so that they can use other methods of design to calculate the specific answer. These typical forms of structure aiming at the specific problem of design is called the prototype. In some degree, the prototype describes the specific structure of the attribute space. This structure adopting the prototype to explore the attribute space and get the intelligent design is called Prototype-Based Design, which is called PBD for short.

Constraint-Satisfied Design (CSD) Method

Constraint-Satisfied Design, which is called CSD for short, regards the design as a Constraint-Satisfied Problem (CSP) and then gets the answer. In Artificial Intelligence, the basic method of CBP is to search for the solution space to find the solution to meet all the problems. CSD usually adopts other intelligent designs to produce a program of design and then estimates whether there is a constraint which satisfies the problem of design. But the simple searching method is usually used in the sub design problem. Because the content of design constraint is various, there are many forms of appearance about it. The most common one is the presentation knowledge of predicate logical form, but there are also many constraints with the precondition. Thus, the constraint includes the precondition and constraint content, and it has a form which is similar to the rules. In addition, there are some complex constraints which have corresponding special methods.

Knowledge Based Reasoning Method

The knowledge program method of CAD is the one that abstracts the experience of designers and shows it as the knowledge, then under the direction of knowledge, through studying the knowledge, it can modify the database and improve the design ability of system. Because ICAD is related with the knowledge inevitably, so ICAD is called “knowledge-based design system”. The typical method of knowledge program is the expert system. A basic model of a expert system is: database with reasoning machine. The knowledge database of design problem can be divided into two kinds:

1. **Knowledge of design process.** It is about the knowledge of how to do the research, including the common principle and experience of experts
2. **Knowledge of design object.** It is about the components, structure, materials and use, and it can also be proto-structure and type of components of some typical product.

The description of the design object is the key problem for the expert system to solve. The feature of design is the final object of design is clear up in the process of design gradually. There are many ICAD systems realized by the expert system. It can be divided into several kinds on the method of knowledge expression: production system, framework and system based on the assumption that the truth maintenance system (ATMS).

CASE STUDY

This section presents an intelligent CAD development platform based on KMCAD system, ac-

cording to special requirement, special intelligent CAD system for diffient project design can be developped rapidly based on this platform.

Background and Motivation

It needs to re-draw pictures in order to renew or modify the design result when using traditional two-dimensional traditional CAD system, especially in designing multi-view part. When modifying the design, there is no direct relationship between the expression of the part and its relative design parameter and they can't be put together, so it's troublesome to save and renew these technical information. In AutoCAD, there is easier way to modify two-dimensional picture, but it is to modify the expression line not the design conception, therefore, this way is still troublesome and quite unreliable.

However, the three-dimensional CAD system demands more knowledge and operational skill and the user have difficulty to master it.

If there was a kind of CAD system by which the designer only needed to input the parameter values and the system produced the assembling view, partial view and other design documents automatically, the designers wouldn't draw pictures and it would be more convenient to operate the system and improve the design efficiency greatly. Therefore, it would make great sense to research and develop such a kind of system.

1. **Designers can be free from the time-consuming and toilful geometric modeling.** With the help of CAD system in this section, designers only need to enter parameter values simply,without drawing any graphics, all the graphics work is completed by the computer.
2. **The design ability of traditional CAD system is improved.** With the help of the knowledge reasoning technology being introduced intoCAD system, the system knowledge processing capability is enhanced, part of the designers' work is replaced effectively. As a result, both product design time and designs' work intensity are reduced significantly, then the competitive ability of enterprise product is improved.
3. **The method to integrate general CAD system with special CAD system is studied.** With this method, the special CAD system is implemented, which can reduce system development time and improve stability of system.
4. **A product design-oriented intelligent CAD system can be exploited in this platform.** All the public modules are encapsulated. For different product design, it only need to develop knowledge module and drawing module, which can shorten the development cycle of CAD system significantly.

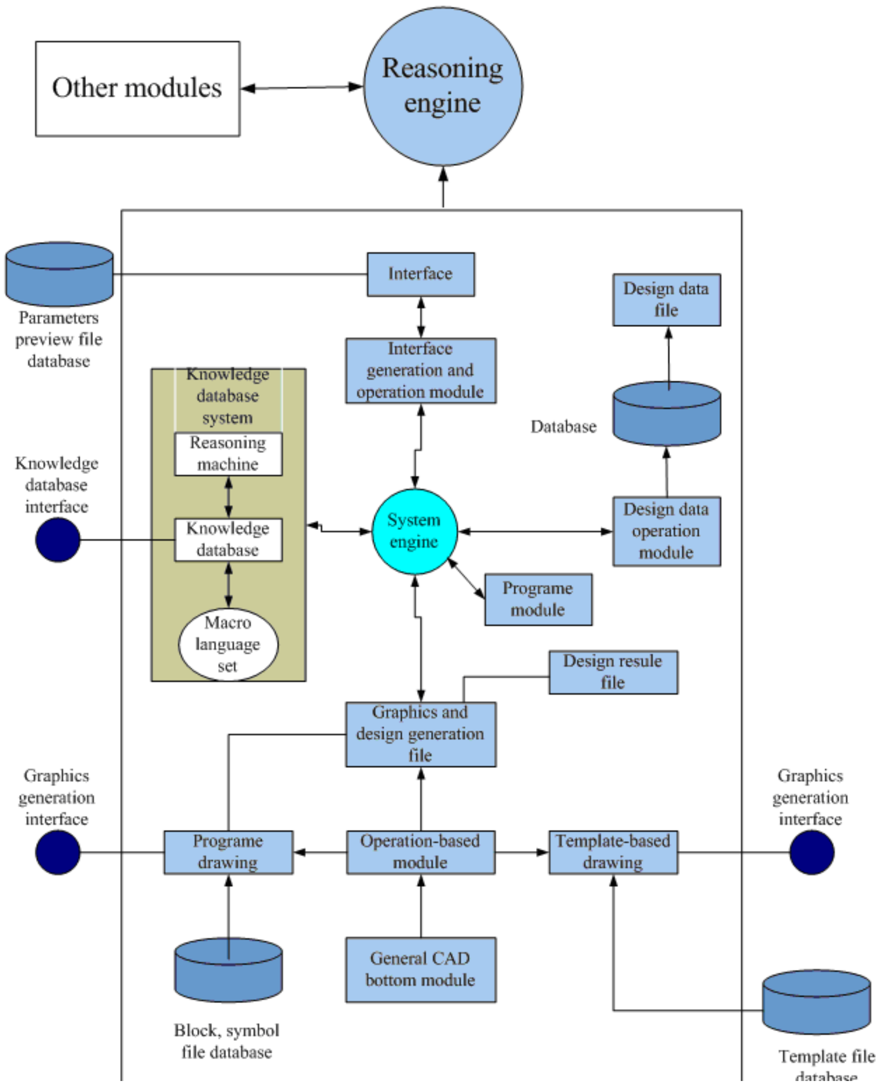
The main idea of the full-parameterized intelligent design technology is to predefine some parameters according to the product design. The designer will be guided into the right parameter-input interface by the system according the design knowledge and only input the corresponding parameter values, then the system will produce all needed engineering pictures and design documents. In the whole process, the designer don't need to draw any pictures.

Modeling

The Overall Design of the Development Platform

The platform is the second development on the KMCAD platform in order to shorten the development cycle and make the intelligent CAD system developed on the platform produce the picture file that can be identified by KMCAD system. The system uses not only the basic picture operating interface of the KMCAD platform but also its printing center to carry out batches of picture printing. Besides, the other types of files can be

Figure 4. Structure of platform of intelligent CAD development based on parametric design



converted into the files in identified format through converting center.

The platform consists of program frame module, knowledge database,database, interface, drawing program, system engine and so on. The relationship of the modules is showed in Figure 4. The platform is built by star structure in which the database module, drawing module, interface module and knowledge database module are completely unattached. The system engine module depends on the above four modules and the program frame module depends on the system engine module, by which the interface element handle can be passed to the interface module. The platform provides three development interface. The needed intelligent CAD system can be developed by adding corresponding content through the three interfaces when developing the intelligent CAD system on the platform. The product design knowledge in required format can be added to the knowledge database by the development interface of the knowledge database system. The develop-

ment of the drawing program can be carried out by the two picture producing development interfaces . The template file database provides the modeling documents with defined parameter variables used when producing pictures. The parameter preview and documentation database provide the picture files needed in parameter preview and save the files with parameter explanation.

Design of Knowledge Database System Module

The intelligentized design based on knowledge of the system can be reflected by the following aspects:

1. **The system has expert design ability.** When using the system to design product, the system can get the design data automatically based on the requirement. Also, the other related design data can be deduced according to the already input design data instead of consulting the product design handbook.
2. **The system can control the design process.** The system can automatically determine the next design process.
3. **The system can automatically verify the rationality of the design result.** After completing the current process, the system first verifies if all the parameters' value accords with the design requirement. The system won't enter the next process even if there is only one parameter value not coincident with the design, but to remind the user to modify the parameter value until all the parameters' value of the process accords with the requirement.

The knowledge database system is built by the structure of "knowledge database +reasoning machine", in which the knowledge database is the file of product knowledge saved by a certain format.. These knowledge includes product explanation and the experience of the product design. The reasoning machine is the mechanism using these knowledge to reason. When the knowledge database is run, the files in knowledge database are compiled and carried out and the knowledge in knowledge database is read into the database in EMS by system engine. When reasoning, the system engine first finds the corresponding formula knowledge in knowledge database through the knowledge in EMS database and then the reasoning maching carries out the reasoning. According to this design idea, the knowledge database system is developed as Figure 5.

The knowledge database system includes three kinds of logic reasoning rule. The method to implement the system is to encapsulate two reasoning functions separately according to each parameter, one function is used to achieve automatic calculation of parameter values, the other is used to achieve automatic validation of parameter design result. Meanwhile, each design flow has an encapsulation with a flow control function. In order to enable the system to call these functions for reasoning, the knowledge database needs to include the function name corresponding to each function. The knowledge in knowledge database is shown as the data structure in Figure 6.

According to structure characteristics of knowledge database, the reasoning machine in this system achieve the reasoning mainly through a function. The function takes character string as parameter, its expression is:

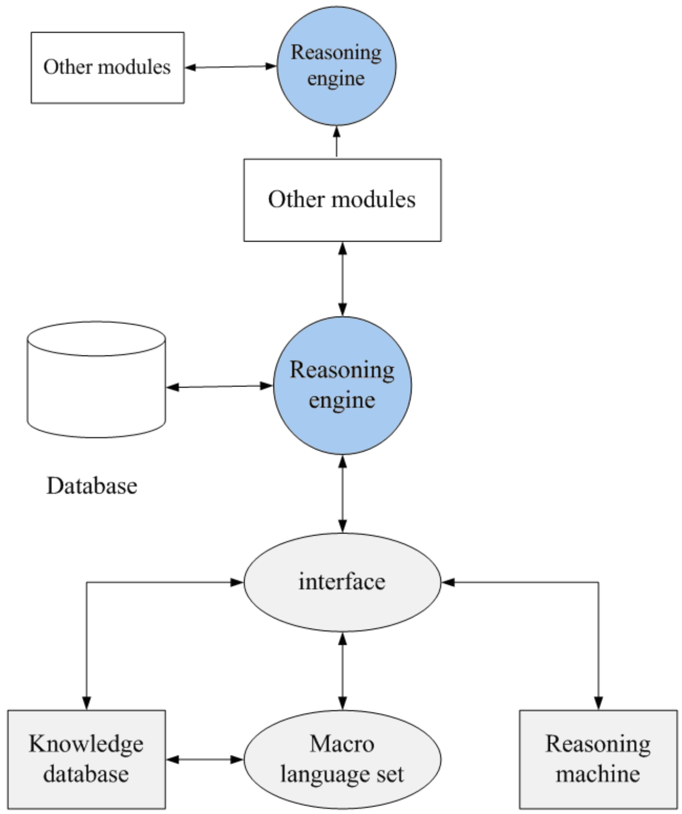
BOOL Reason_Kowledge(CString string)

The system achieves reasoning by running reasoning machine, consequently, intelligent design is realized.

Graphics Producing Module Based on Parametric Design

With the intelligent CAD system in this section, designers do not need to draw graphics, all the graphics can be produced by a special module, which is graphics producing module based on parametric design.

Figure 5. Structure of knowledge database system



It is very important to design a structured graphics producing module, the final design should meet the following requirements:

1. Separating with basic operation function of graphics generation, that is, the public functions used in the process of graphics drawing are encapsulated in a module, while the graphics producing module is only responsible for generating all of the related design documents.
2. The part of graphics drawing is separated from the part of graphics file generation.
3. Two kinds of graphics drawing methods are adopted, one is programe drawing and the other is template drawing.

According to the requirements above, graphics producing system is implemented as Figure 7.

Basic operation module: the basic graphics operation functions is provided in this part, including drawing functions, dimension functions, file operation functions, block operation functions, and so on.

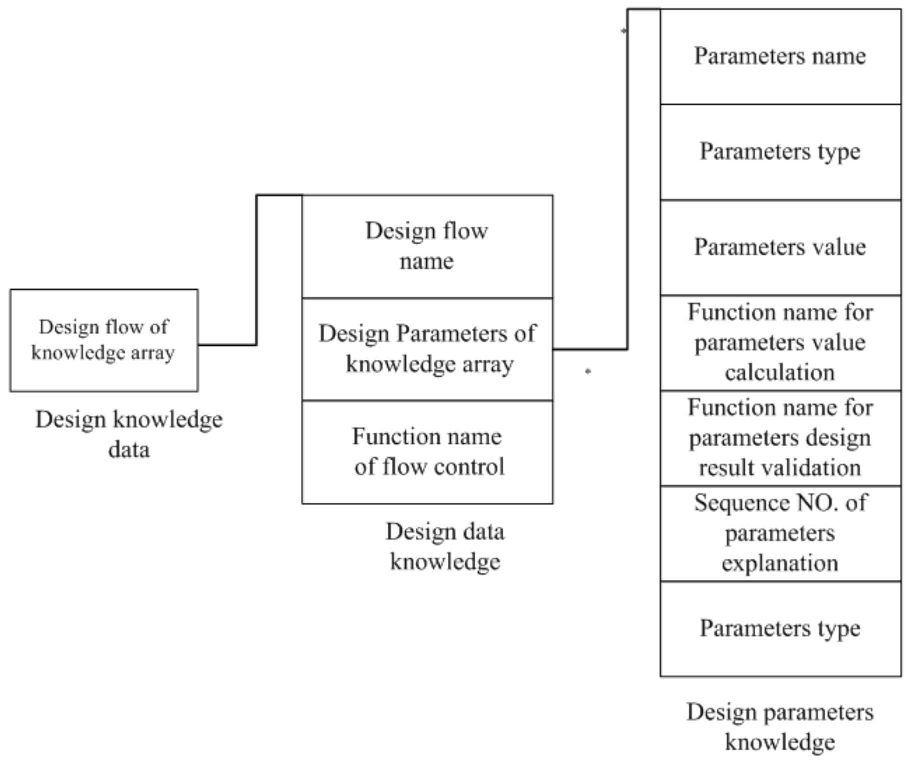
Template-based drawing: the template file is made a small number of modifications to complete the generation of project graphics.

Programe drawing: the generation of a vast majority of information of a project graphics is completed throgh programe, including drawing graphics, marking parameters, filling out the list, and so on.

Conclusion

As the application of CAD software is deepened, the deficiency of the traditional CAD system has been exposed gradually. First, a large amount of

Figure 6. Data structure of knowledge database

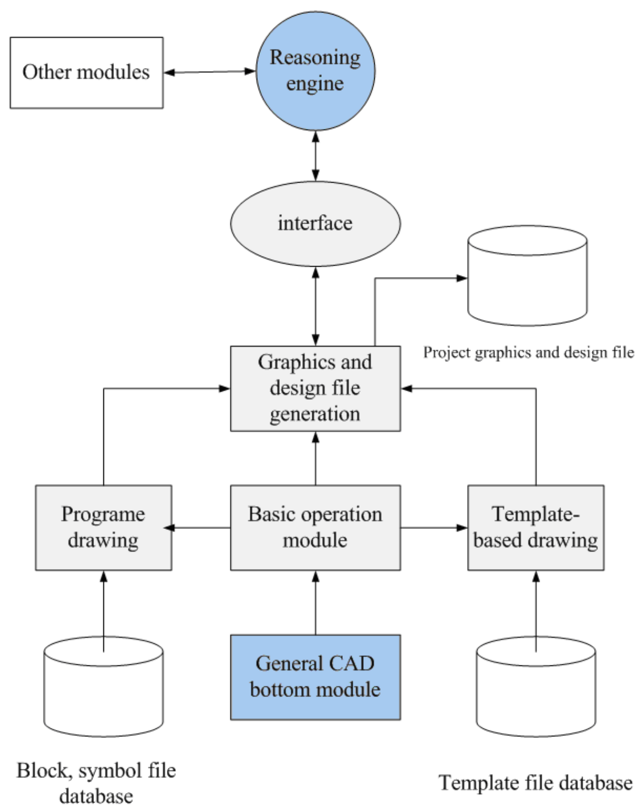


work is done repeatedly in the design process of the similar product; on the other hand, the support to design is insufficient in the design process of products.

In order to solve the above problems, this thesis researches and develops an intellectual CAD technology and system based on parametric design which is used for developing intellectual CAD system for the design of a specific product. This intellectual CAD system not only has ability of intelligent design but also has ability to produce the figure automatically. The former can provide support for designing products; and the latter can free the designer from the tedious repeat drafting operations. The function of intelligent design can be realized through using the knowledge base system which is structured on "knowledge base+reasoning machine" and carry on reasoning via knowledge reasoning technology for intelligent design. The knowledge base is comprised of abstract products design knowledge, and the reasoning machine is a mechanism carrying on reasoning in terms of the knowledge in the knowledge base. Function of producing the figure automatically is realized by the special system which can produce all figures with two kinds of methods to produce figure, one of which is that figures are produced by revising templates, and the other is that figures are drawn through the program.

In order to shorten the development cycle of intellectual CAD system, a developing platform has been set up. Based on the platform, an intellectual CAD system for specific product design can be developed rapidly. The development process can be divided into two steps. Firstly, summarizes the design knowledge of the products according to the user's requirements and adds them to the knowledge base. Then finishes development of

Figure 7. Structure of graphics producing system



the drawing generation system which can meet user's demand perfectly.

SUMMARY

I^2CAD technology is the trend and focus of CAD technology, however, it is a comprehensive and complex systems engineering. In order to design a perfect and high efficient I^2CAD system, it is necessary to combine the experts in various fields and the expertise and technology for product design service, then improving the product design efficiency and quality to lay a solid foundation for enhancing the product and market competitiveness of enterprises.

In the 21st century, the basic features of ICAD technology is high integration, intelligenization and standardization. The development targets of ICAD are enhancing product quality and manufacturing efficiency, shortening designing and manufacturing cycle, reducing manufacturing costs, improving reactivity as possible to meet the requirements of users.

By bringing artificial intelligence into the CAD technology, the computer may be used for product innovation directly which can maximally improve human being's creativity. Then bringing multi-media technology, network technology into the CAD field and synthesizing the cooperation function between designer and computer greatly help the enterprise renovating of renewing products and capturing market. In the next 10 years, the development direction of CAD technology can be called I^4CAD, namely, Integrated, Intelligent, Internetworking and Interactive CAD.

The development trend of ICAD is expected as follows.

Improving the Designing Ability of Traditional CAD

The knowledge reasoning technology is adapted in CAD system to enhance the ability of dealing with knowledge and replace part working of designer effectively. Thus design time of product is shortened greatly and the working strength is decreased to increase the enterprise competitiveness.

Researching the Way of Integrating Common ICAD with Special ICAD

For the advanced efficiency of product design, special ICAD facing on enterprise (or product) based on common ICAD system is necessary. It utilizes the technology of exist ICAD and have the advanced ability of product design.

Developing More Excellent ICAD System Platform for Product Design

Combined other technology such as knowledge-base system, multi-agent, data mining and knowledge discovering, intelligent computer, multi-sensor integration and so on with CAD, the new ICAD system is desired to be more intelligent, integrated and interactive. For different product design, only the product design knowledge module and plan operation module need to be developed, so the developing of CAD system is effective.

REFERENCES

Bellifemine, F., Poggi, A., & Rimassa, G. (2000). Developing multi-agent systems with JADE. In proc. of the 7th Intl. Workshop ATAL 2000, (pp. 85-99).

Bussmann, S., & Jennings, N. R. Wooldridge. (2004). Multi-agent Systems for Manufacturing Control - A Design Methodology. Berlin: Springer.

Coad, P., & Yourdon, E. (1998). *Object Oriented Analysis, Object Oriented Design*. Upper Saddle River, NJ: Yourdon Press, Prentice Hall.

Douglass, B. P. (1999). *Real-Time UML* (2nd ed.). Reading, MA: Addison Wesley.

Gomaa, H. (2000). *Designing Concurrent, Distributed, and Real-Time Applications with UML*. Reading, MA: Addison Wesley.

Jia, H. W., Tang, W. Q., He, T., et al. (2005). A Conflict Resolution Methodology Based on Technology and Sociology- A Complex-System Approach. In *Proceedings of the Ninth International Conference on CSCW in Design, May 2005*, (pp.16-22).

Kayır, Y. & Gülesin, M. (1996). Makina ve İmalat Mühendisliğinde Uzman Sistemler. *Makina Market,* (3),100-104

Knudsen, S. U., Pedersen, T. B., & Thomsen, C. (2005). RELAXML: Bidirectional Transfer between Relational and XML Data. In *Proceedings 9th International Database Engineering & Application Symposium,* July 2005, (pp. 151-162).

Lin, P. (2006). Intelligent CAD Expert System Platform Development and Realization. *Computer Engineering and Design*, *27*(12), 2151–2153.

Nathan, D. M. (1998). *An Integrated Environment for Conceptual Design, Synthesis and Analysis of Dynamic Frame Structures*. Doctoral Dissertation of the University of Utah.

Ouyang, M. A., Li, C. G., Zhong, Y. F., Yu, J., & Zhou, J. (1996), Intelligent CAD Approach For Modular Design. In *Proceedings of SPIE - The International Society for Optical Engineering*, (Vol. 2644, pp. 614-619).

Paszek, R. & Knosala. (1997). The Method of the Knowledge Representation In an Expert System for Metal Cutting Engineering. *Journal of Materials Processing Technology*, *64*, 319–326. doi:10.1016/S0924-0136(96)02582-4

Rao, A. S., & Georgeff, M. P. (1995). BDI Agents: From Theory to Practice. In *Proc. of International Conference on Multi-Agent Systems (ICMAS),* San Francisco.

Rosenberg, D., & Scott, K. (1999). *Use Case Driven Object Modeling with UML*. Reading, MA: Addison Wesley.

Rumbaugh, J., Booch, G., & Jacobson, I. (1999). *The Unified Modeling Language Reference Manual*. Reading, MA: Addison Wesley.

Shen, W., & Barthe`s, J.P..(1997). An experimental environment for exchanging engineering design knowledge by cognitive agents. In M. Mantyla, S. Finger, & T. Tomiyama (Eds.), *Knowledge Intensive CAD-2* (pp. 19–38). Boca Raton, FL: Chapman &Hall.

Toye, G., Cutkosky, M.R., Leifer, L., Tenenbaum, J., Glicksman, J. (1993). SHARE: A methodology and environment for collaborative product development. In *Proceedings of Second Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, (pp. 33–47)

Xin, T., & Ray, I. (2005). Detection for Conflicts of Dependencies in Advanced Transaction Models. In *Proceedings 9th International Database Engineering & Application Symposium*, July 2005, (pp.151-162).

Y. L. (2000). *Integrated Infrastructures for Collaborative Development in a Virtual Enterprise*. PhD Thesis, Tsinghua University, China.

Chapter 11
Intelligent Process Planning:
Intelligent CAPP

Computer Aided Process Planning is very hot topic in the manufacturing. It uses the geometric information (such as shape, size, etc.) and information technology (such as materials, heat treatment, bulk, etc.) which are input into the computer to output parts of the route of the process and the procedures automatically. Process planning is very important in the manufacturing process. With the continuous development of the manufacturing sector, the traditional manual methods of Process Planning flaws more and more serious. Computer-aided technology can increase their technical capacity effectively. CAPP is an effective means to improve the design. The research of CAPP has got a very huge development, from the search logic structure, Variant, Generative, and Hybrid to Expert System. In the future, the development of the CAPP will focus on the extending of the application scope, depth and level. In this chapter, a general introduction is presented firstly. Then the application of genetic algorithm (GA) to CAPP is introduced. Thirdly implement of ANN in CAPP System is presented. In the fourth part, use of Case-Based Reasoning in CAPP is discussed. Fourthly, CAPP based on Multi-Agent (MAS) system is illustrated.

DOI: 10.4018/978-1-60566-864-2.ch011

INTRODUCTION

Computer Aided Process Planning is very hot topic in the manufacturing. It uses the geometric information (such as shape, size, etc.) and information technology (such as materials, heat treatment, bulk, etc.) which are input into the computer to output parts of the route of the process and the procedures automatically. CAPP uses computer to design and build a bridge between the manufacture and the design. It belongs to the field of engineering analysis and design, and it also stressed that the automatic design process.

The Origin of CAPP

Process planning translates design information into the process steps and instructions to efficiently and effectively manufacture products. Process planning is very important in the manufacturing process. It is the interface of product design and manufactur-

ing, in this process, vast amounts of information must be analysis.

Process planning is affected by a lot of factors such as the production environment, product type, manufacturer resources, production volume, process methods, the level of the experience, and even restricted by the production organization and management. These factors above in any changes would lead to the change of the programmer design, so that the process is full of "personality."

The process planning outputs a large amount of information, which guide man to the procurement of enterprise material, the production planning and scheduling, the organization of production, the balance of resources, the cost accounting and so on. It encompasses the activities and functions to prepare a detailed set of plans and instructions to produce a part (Chang T. C., Richard A. & Wysk, Wang H. P., 1998). The planning begins with engineering drawings, specifications, parts or material lists and a forecast of demand. The results of Process planning are the following:

Routing Design

Routings specify operation sequences, the operations on the station or work centers, the tools and fixtures. And the gotten routing becomes a major input to the manufacturing system, with which to define operations for production activity control purposes and to define required resources for capacity requirements planning purposes.

Process Planning

Process planning typically provides a detailed, step-by-step work instructions, including the individual operations, machining parameters, set-up instructions, and quality assurance checkpoints.

With the continuous development of the manufacturing sector, the traditional manual methods of Process Planning flaws more and more outstanding. There are as follows:

1. There is a lot of duplicate work. The data is double counting and unable to use the CAD information directly, so it is a heavy workload, and low efficiency, long cycle, error-prone.
2. The quality of process planning depends entirely on the technological level and work experience of the staff; it is difficult to guarantee the accuracy of data.
3. The methods of technology vary from person to person, it is difficult to achieve the succession process design, regulatory, standardization and optimization, and cannot reduce production costs effectively.
4. It is impossible to use the computer to achieve unified management of process documents. The information of the process planning cannot be shared with the other departments. The technology and the knowledge cannot be made full use; neither do the manufacturing resources and the technology resources.
5. The technician must have rich experience in production, and be familiar with various internal processing methods and the use of appropriate equipment. Also it means that as a good technician he must be familiar with the production and processing enterprises of various regulations.

The increasingly fierce competition and the increasingly volatile market require enterprises continuously introduce new products, to make rapid response to their customer. At the same time, it requires more technology and shorter delivery time. All of this is severe challenges to the technical capacity of enterprises, technological capacity of our company.

Computer-aided technology can increase their technical capacity effectively; CAPP is an effective means to improve the design. The implementation of CAPP system, promote enterprise information technology departments is an important subject to enterprise.

The Process of CAPP Development

The study of Computer-aided design started in the late 1960s. The early intent was to establish a computer-aided system which included the generation of the process card, the storage of the process content and the search of the process. It used the computer to store, organize, compute and extract information, just a tool to help the staff to reduce the businesslike work, so that the time of design is saved. These systems are not universal because they do not have the decision-making ability and the sort function. And the true CAPP system which has the universal significance was designed in Norway calls AUTOPROS in 1969, and then a lot of the CAPP systems are subject to the impact of this system. The CAM-I'S automated process planning system, which was developed by the U.S. computer-aided manufacturing international organizations CAM-I(Computer Aided Manufacturing-International) as a milestone in the computer-aided process history, named the Computer-aided design process as CAPP.

It is a relatively long and tortuous process of the development of CAPP system. At 1965, Niebel brought forward the CAPP for the first time. From then on, the research of CAPP has got a very huge development, from the search logic structure, Variant, Generative, and Hybrid to Expert System. Now there is a large number of CAPP prototype system and commercialization CAPP system.

The search logic structure CAPP system uses an interactive human-machine dialogue, which is based on standard strip and typical process of design, to design the process planning. The quality has a great dependence on technology.

The Variant Group CAPP system use group technology to classify the targets by comparability(for example, the spare parts or the process similarity; their structure and function of the components, and so on).Then the CAPP system designs the process planning for each group (Cluster) object typical design process, and finally establishes the typical process.

The Generative CAPP system is designed by the logic of the decision-making process and algorithm. It is an automatically generated process. The Generative CAPP system does not need the interrupt of technology, so the consistency is assured. At the same time, because of the variability and the complexity in manufacturing environment, it is v extremely difficult to achieve a CAPP system when the structure is complex and the components are too much.

Hybrid has the excellence of all the above. So a large number of CAPP systems take this model now.

It is that concept brings the method of using the Tooled-CAPP, its main idea is:

1. CAPP is a tool to free the process design staff from the fussy design work;
2. Automation is not the only target of CAPP;
3. The main objective of CAPP system is to achieve the easy use of operation, the efficient means of process making, the automatic aggregate statistics of the process information, the information integration with CAD / PDM / ERP systems. Having a good open and integrated port is very important for the research and extension of the CAPP system.

The Tooled-CAPP is a great success to the commercial of CAPP system. It makes the CAPP develop from the laboratory to the market and enterprises. With the Tooled-CAPP system, thousands of enterprises improve the efficiency of the process design. And it also promotes the standardization of this technology and the building of Enterprise Information.

The Expert CAPP system bases on the artificial intelligence. It is an automatically generated process, just like the Generative CAPP system. But the Generative CAPP system depends on

the logic of the decision-making process and algorithm, when the Expert CAPP system is on the repository and inference engine.

The Trend of CAPP Development

CAPP has been developed for many years, and holds a certain market, but it never enters to the period of the market outbreak, and also have not been widespread promotion and application. This cannot be separate from the deficiencies and problems in the current CAPP application:

1. The big deviation between the research and the development;
2. The inadequate attention on the systematic, practical and engineering study of CAPP;
3. The lack of the basic technology research; the integration between CAPP system and other systems.

Throughout the development course of the CAPP System, the research and application of the CAPP has always circumfused on two aspects of needs: On one hand, they constantly improve their own shortage during the application; on the other hand, they constantly meet the new requirements brought out by the new technology and manufacturing mode. Therefore, the future development of the CAPP will focus on the extending of the application scope, depth and level. And the CAPP system may have several development trends as follow:

Knowledgeable and Intelligent

CAPP system will not stop on the stage of focus on resolving transactional work. The Expert CAPP system will not only be an instrument of process planning, but also gather the experience and knowledge of technology experts. CAPP system should pay attention to the process design, level of physical characteristics or provide options in the whole process of the programmer, and be self-learning, adaptive in accordance with the work record of the operator.

Tool Kited

The technology environment and management models vary. It is not only necessary to adapt to the specific situation of each enterprise, but also to improve the versatility and the workload of the implementation. The CAPP system should be divided into a lot of dependent toolkit, and then the users could assemble the toolkit for their own companies.

Integrated and Networked

CAPP is the bridge between CAD and CAM; the important sources of information for CAQ, PDM and ERP, and it also need the product design feature information provided by the CAD model. The development of these systems can be relatively uneven (Yi H. & Ni Z.H., 2003). CAPP and CAD, CAM (e.g.) must be fully integrated under the thinking of the parallel project .Only in this situation can CAPP play as the information hub and functional regulation in the whole production activities. For example, to exchange in two-way and send the information with product design.

Networked is a requirement for integration application in modern network system. The design of the role and the type in parallel process, the exchange and send with CAD, and the integration with CAQ, CAM, PDM all need network to support.

Interactive and progressive CAPP system is used to help rather than replace the process designer, so the common tooled-CAPP system should not be pursued fully automated. Operators should have sufficient technical knowledge and judgment ability since the key decision-making should be made by the operators. Decision-making and judging is not very difficult work to the technologists who have enough ability of the judging of the techniques, but it may be difficult

for the computer to complete. Knowledge Base and its use rules need to be gradually established, validated and perfected. And the commercial and knowledge-based CAPP System needs to be developed with goals and plans gradually.

Now CAPP, as an important model in CIMS, should be more knowledgeable, intelligent, tool kited, integrated, and networked. CAPP should make greater contributions to the informatization of enterprises.

APPLICATION OF GA TO COMPUTER-AIDED PROCESS PLANNING

A dynamic workshop environment exhibits a high degree of complexity where the activities of sequencing machining, selecting machining resources and obtaining an optimal or near-optimal process plan should be considered simultaneously to achieve the highest production efficiency (Zhang F, Zhang Y.F. & Nee AYC, 1997). Process planning determines the detailed requirements for transforming a raw material into a completed part, with the available resources of workshop .Therefore, developing a CAPP system, which can produce optimal process plans in a dynamic workshop environment, is important.

Description of Sorting Problem Based on GA

The sorting problem of process is one that optimizes the sorting of component characteristic process, it is not to obtain the optimization value but sorting optimization. As for the optimization problem of obtaining the function maximum, generally it is described by following mathematics model:

$$\begin{cases} \max & F=f(X) \\ \text{s.t.} & X \in R \\ & R \subseteq \Omega . \end{cases}$$

In this formula, $X=[x_1,x_2,\ldots,x_n]^T$ is decision independent variable, it can be either number or logic variable; f(x) is object function; Ω is the basic space of problem, R is a subset of Ω, the solution X that satisfies the restriction is called practical solution. R is composed of all the solutions that satisfy the restriction and it is called practical solution. The optimization object is to find a solution $X_o(x_{1o},x_{2o},\ldots,x_{no}) \in R$, which can make sure that $F=f(x_{1o},x_{2o},\ldots,x_{no})$ obtains the maximum(or minimum, or other objects).

As for the optimization problem mentioned above, there are various object functions and restrictions, some are linear, some are nonlinear, some are sequential, some are discrete, some are single peak,and some are multi-peak. With the advancement of research, gradually, it is impossible and impractical to obtain the exact optimization solution in many complicated occasions, thus, it becomes the main object to obtain the approximate solution or the optimization solution. GA provides us with an effective method and commonly-used frame, and starts up a completely new optimization algorithm.

In GA, $X=[x_1,x_2,\ldots,x_n]^T$with N is noted as follows.

$$X=X_1X_2\cdots X_n \Leftrightarrow X=[x_1,x_2,\ldots,x_n]^T$$

Each Xi is viewed as one genetic gene, in this way, X can be seen as a chromosome that is made up of N genetic genes. The decision variable X makes solution space of problems, and the search for optimization solution of problem is actually the search for chromosome X, thus, all the chromosome X make the search space for problem.

Biological evolvement takes population as its main body, and its process is completed through crossover and mutation of chromosomes .The search process of optimization in GA imitates the biological evolvement .The operation objects is the complex that is made up of individuals, which number is M. Being similar to biological evolve-

ment from one generation to another, GA is also a iterative process. The t generation is marked as P(t), After one generation's evolvement, the t+1 generation population is obtained, marked as P(t+1). According to the principle of "survival of the fittest" and after continuous evolvement, finally an excellent individual X is obtained, which corresponding X equals to or approximately equals to the optimization X* of problem.

Process sorting is the process decision and characters after making sure of all the process characteristics and operation methods of components. Since the operation chain of component characteristic has been ensured, the operation chain is divided into characteristic operation units, for instance, the characteristic operation chain of one plane M is: "plane+rough milling+fine milling,that is, it is divided into three units: "plane and rough milling and fine milling", the complex of all the characteristic operation units of components is denoted as: F={f1,f2,……, fn}.

The characteristic operation sequence in F is random, and it is not the final operation sequence, generally, F also includes heat operation and assistant operation. Process sequence have to satisfy a set of restrictions, and the restriction complex is marked as: R = {r1,r2,……, rn}.

There exist types of relationships i.e. constrains among different process features of accessories. The process sorting is therefore to work out the mappings in F for every single element of the constriction set R. And an operation sequence is considered to be optimum in case that it meets the constrictions far as possible.

The optimization of process sorting varies from the normal. For one thing, the objects that are optimized are not numerical data but sequences. For another, it is difficult to express the explicit optimize index using fitness functions. Thus the constrains and fitness functions are to be analyzed and then designed for the control strategies of global search when the GA method is applied to calculate the optimum operation sequences.

Given that Ω is a set of operation sequences, the calculation of GA can be described as follows: In the domain of Ω, the maximum value of fitness function for the set of processing features St = {ft1,ft2,……ftn} is the optimum.

Given that Ra is the set of compulsory restrains, and Rb the set of optimized constrains.

In the process practice, Ra has an impact on the sequence, which is of the most importance of all restrains. Relationships that must be considered include:

1. The datum is of the highest priority,
2. Moulding surfaces are of higher priority than the holes and key seats, and
3. Main features, which define an accessory at the basic level, are of higher priority than the sub features, which define an accessory in greater detail.
4. Non-destructive constrained. Any operations cannot destruct features that are implemented by operations beforehand i.e. the inner surface of a cylinder must be implemented before the screw thread.
5. Accessibility constrained. Some features can only be implemented after some others so as to ensure that every single feature is accessible.
6. The sequence of features implementation is restrained by properties of features themselves in some cases. For instance, needless to say that the process of hole enlargement should be operated after drilling.

In terms of the consumed time, accuracy, and cost-effectiveness, Rb in most cases include:

1. Reduce the number of Times for clamping as far as possible. It does take time to clamp however clamping is in need throughout the processing.

Given that the set of processing features is {f1f2f3……fn}, and the set of clamping

types {SU1,SU2,…,SUk},(k≤n). For SUi (i ≤ k), the set of features that can be processed is SUi={fi1,fi2,…, fim}, and clamping must be carried out if the next feature is not in SUi. Apparently, given that Ns is the number of times for clamping, MIN(Ns)=k and MAX(Ns)=n. That is, in the optimum circumstances, all features of the same type are processed after the corresponding clamping is carried out, and vice versa.

2. Reduce the number of times for tools' changing. Given that the set of tools that are used for processing one accessory is {T1,T2,… Tl}, (l≤n) in which the set of features that can be processed by tool Ti(i≤l) is Ti = { fi1,fi2,…,fim }. The tool must be changed after processing some feature in Ti if the next feature is not in Ti. Given that Nt is the number of times for changing tools, MIN(Nt)=l and MAX(Nt)=n. That is, in the optimum circumstances, one tool is used for processing all features of the same type, and in the worst circumstances, tools are changed after the processing of every single feature.

Genetic Algorithm on CAPP

In GA, the signification of a sequence of date known as chromosome or string is the sequence of the operations or a candidate plan, further to say, in a distributed manufacturing system, a chromosome not only represents the sequence of the operations but also indicates which factory the operates come from.. The chromosome is composed of several genes which include factory ID, operation ID, corresponding machine, tool and tool access direction (TAD), and so on. Table 1 shows the representation of a five-operation process plan. '01' is the factory ID, 'Op4' represents operation 4; m-01, t-05 and +y in the second row represent the machine, tool and TAD, so are the other columns.

Table 1. Representation of a process plan

01	Op4	0p1	Op3	Op2	Op5
	m-02 t-05 +y	m-03 t-02 +x	m-02 t-04 +z	m-01 t-04 -x	m-03 t-03 +y

The genetic algorithm includes four main operates is respectively population initialization, reproduction, crossover and mutation. Based on the each chromosome's fitness degree to select candidate chromosome, then complete reproduction, genes recombination, and mutation.

Population Initialization

The generation of the initial population in GA is usually done randomly. However, since precedence relations between operations are used, the initial population must consist of strings of valid sequences, satisfying all precedence relations. The rule to generate the initial population is represented as follows: select one operation among the all feasible operations, which have no predecessors or which either predecessor all have already been selected. Repeat select until each operation has been selected for only once. Revisit the first selected operation. Randomly select machines and tools from the selected factory that can be used for manufacturing the operation, synchronously, select one amongst all possible TADs for the operation, until each operation has been assigned a machine, tool and TAD. Repeat the all methods above, until the number of chromosome reaches prescribed number.

Reproduction

The "roulette wheel" method is used for the reproduction of the strings: (1) It's to calculate the summationof all chromosomes' fitness value.(2) This method leads to find a random number (m) between the summation and 0.(3)Begin with the first chromosome in population. Calculate the

summation(A) of the first and the next chromosome's fitness until $A \geq m$.In(3),the last chromosome is selected to reproduce.

Because of that the realization of reproduction based on probability, the chromosome with higher fitness value has more chance to reproduce than the one with lower fitness value. Chromosome with low fitness means that it has few copies (or no copy) in new generation. AS a result, reproduction will improve the average fitness value of the next population. But on the other hand, a disadvantage brought by reproduction is the chromosome whose fitness value has mutated will appear in the next population, it'll lead to the standstill of search, even lead to a wrong result. In GA, the algorithm is improved by crossover and mutation.

Crossover

The crossover operator is applied, at a given probability, to the strings that resulted from reproduction.

The crossover operator must ensure that precedence relations are maintained and that only feasible strings are generated.

Randomly choose two chromosomes as parent chromosomes. Based on the chromosome length, two crossover points are randomly generated to select a segment in one parent. Each string is then divided into three parts, the left side, the middle segment and the right side according to the cutting points. Copy the left side and right side of parent 1 to form the left side and right side of child 1. According to the order of operations in parent 2, the operator constructs the middle segment of child 1 with operations of parent 2, whose IDs are the same as operations of the middle segment in parent 1.The role of these parents will then be exchanged in order to generate another offspring child 2.Re-assign machines and tools to the operations in the middle segment to legalize the offspring chromosomes according to the factory ID.

Mutation

The mutation operator is applied with a small probability to introduce some genetic diversity, thus avoiding being struck at a local optimum. Once again it is important that the application of the mutation operator generates only feasible strings. In order to avoid the newly generated chromosome may be invalid because of the violation of precedence relations, an algorithm could be applied to the mutated chromosomes guarantee its feasibility. The pseudo codes of the algorithm are:

Begin

A=1

While A less than (subtract1from the number of genes)

B=A+1

While B less than (subtract1from the number of genes)

C=0

if C less than number of successors of chromosome(B)

if chromosome (A) equal to successor (C) of chromosome(B)

Exchange chromosome (A) and chromosome (B)

C++

B++

A++

End

Three mutation operations were used. The first one refers to the string and consists on the random exchange of two genes of the string with a predetermined probability. A feasibility test is applied in order to guarantee that the string is feasible and the mutation is applied only if this test returns true. The other two mutation operators refer to machines and tools and they are similar. The algorithm is the following: randomly select a position in the string with a predetermined probability and randomly choose a machine (or tool) from all the alternatives to replace the current machine (or tool).

Before reproduction, chromosomes' fitness value have been mentioned, usually two optimization criteria are used, namely minimum processing times and minimum production cost, to calculate the fitness of each process plan and measure the efficiency of a manufacturing system. Processing time is generally comprised with machining time (MT), machine change time (MCT), tool-change time (TCT) and set-up change time (SCT). Different from the other three time indices, machining time index assumed to reflect the importance of machining time in the overall processing time. It is assumed that the machining time of a unit volume is treated as a fixed amount for a particular combination of operation type and machine. User define the machining time indexes for each type of operation combined with machines, Thus, given the removed volume and machine for manufacturing a feature through one single operation, this feature's machining time will be the result of multiplication of MTI and the removed volume.

The other optimization criterion is production cost.

$$SPC = \sum_{i=1}^{n}(OC_i + SC_{i+1} \times \alpha(OMT_i, OMT_{i+1})),$$

SPC is system production cost is the number of operations. OMT is a combination of operation-machine-tool. OC is the machining cost for OMT I and SC is the setup cost for changing for OMT_i to OMT_{i+1}. $\alpha\left(x, y\right) = \{^{1, x \neq y}_{0, x = y}$ .The setup cost is calculated as a sum of following costs: machine setup, tool setup, machine change and tool change. The validities of OMTs are emphasized; one matrix position has a value only if the precedence relations allow the transition between the two corresponding OMTs.

In conclusion, for the complex parts with numerous processing characteristics, among the characteristics, there are many constraints and contradictions,the use of traditional mode becomes difficult. An algorithm for process planning is chosen-GA. This algorithm takes full advantage of global searching and optimization performances of GA to search reasonable process route (Joao Rocha, Carlos Ramos & Zita Vale, 1999).

Compared with Traditional CAPP

In the traditional CAPP system, sequence of process route always is determined by hierarchical constraint. The constraints include geometry shape, technologic method, and optimization index and so on. It's inefficiency to express the technologic decision knowledge and realize the decision logic. This method of obtaining optimal sequence of the working procedures is so difficult that it becomes a bottleneck in CAPP.

Based on genetic algorithms that are introduced to sequence the process, where exists different types of relationships among the process features of complex accessories, which can be categorized as compulsory restrains and optimized constrains according to the requirements. In the operations and steps sequencing, the former is of higher priority, and the latter is expected to be met as far as possible. Thus, the process of optimized sequencing is to find out a set of feasible operation sequences that meet the reasonable constrains, and then select the one that is of higher quality than any others in the set in terms of the optimization standards.

THE IMPLEMENTATION OF ANN IN CAPP SYSTEM

In recent years, almost all kinds of expert system have the defects such as difficulties of KA, lack of self-learning, explaining and associating ability. Owing to this actuality, engineers and experts have made great efforts to bring out lots of new methods and theory. such as object-oriented approach, fuzzy decision-making, Artificial Neural Network, eigenvector extraction and so on. However engineers are comparatively more interested in Artificial Neural Network and there are more production in this aspect.

Artificial Neural Network model is a nonlinear dynamics network system based on the research of modern neurophysiology and psychology. It is up built by limitating the structure of human brain's neurons. It doesn't need any rules or tree searching if the Artificial Neural Network theory is applied to build CAPP system. The machine replaces symbol ratiocination by computation ratiocination via applying the ratiocination mechanism of the Artificial Neural Network model. There are two main aspects about ANN's applying to a CAPP system. On one hand, the Artificial Neural Network is used as a feature classifier to design the CAPP system from the perspective of feature recognition; On the other hand, the CAPP expert system based on the Artificial Neural Network is designed from the perspective of knowledge. These two kinds of applications are introduced respectively in next section.

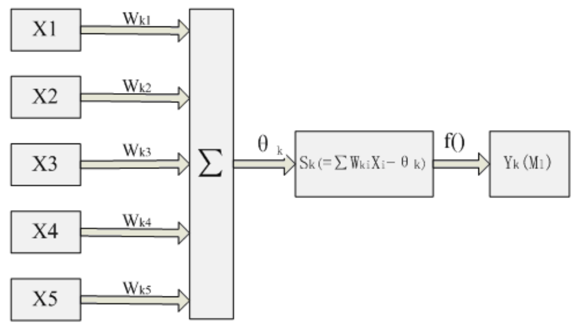
Applying the ANN as the Feature Classifier of the CAPP System

The Establishment of Neuron Model

The core of a CAPP system is techniques process planning, and the basis of techniques process planning is the choosing of processing methods. When choosing the process method for every surface of an accessory, generally about five key factors are concerned, namely the type of surface characteristics, the surface characteristics of the type, feature size, accuracy class of the size, surface roughness, accessory material. If these five factors are expressed as X1,X2,X3,X4,X5, the chain of the processing methods M can be expressed as M= f (X1,X2,X3,X4,X5). If the threshold in the neuron is expressed as θk, the sum output of the neuron as Sk, the weight of every factor as Wki, then the unit neuron model of processing methods choosing can be shown in Figure 1.

This neuron model can be expressed as a formula:

Figure 1. The unit neuron model of processing methods choosing



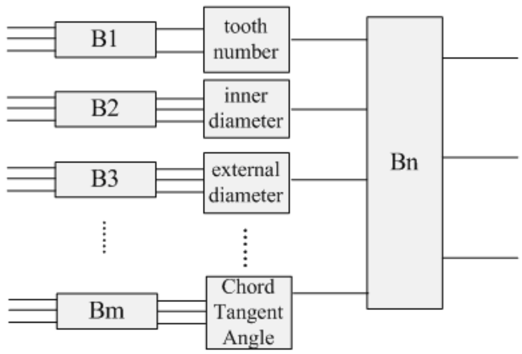
$$Sk=\sum WkiXi-\theta k$$

$$Yk(Ml)=f(Sk)$$

The ANN Structure of the Processing Methods Choosing

The Artificial Neural Network of the processing methods choosing is composed by a number of neurons which are used for choosing processing methods according to a certain topological structure. This network receives all the relevant information about processing methods, forming information masses which including processing methods. BP network is a classic feed forward network which consist of input layer, monolayer and output layer. And the processing method is chosen through the BP network. As a matter of fact, once a single hidden layer has sufficient number of hidden layer unit, it can approach to any finite dimensional measurable function with either precision. This shows that BP network can progress functional approximation thereby complete the nonlinear mapping from accessory characteristic controls to processing controls. With the BP network, the sample's input and output problem into a nonlinear optimization problem is translated firstly, and then the weight of the network is modified by ceaselessly iterative computation using the most common gradient descent algorithm. And finally

Figure 2. The BP network of straight gear



it will implement the minimum MSE between the actual output and the expectation output and get the proper processing scheme (S.Abid, F.Fnaiech & M.Najim, 2000) .

One accessory is composed of some kinds of surface characteristics, and processing methods choosing of each characteristic is a BP network. When a number of BP network combine together, there is one accessory's processing methods choosing ANN (for example, straight gear). The design idea of combined BP network is to decompose the processing methods choosing network of one accessory into a control network and several characteristic sub-networks, as shown in Figure 2. The control network can judge which characteristic sub-network the input characteristics belong to, based on grouping of components' characteristics. The characteristic sub-networks choose the processing methods for the characteristics in each group by using the BP algorithm. The control network and characteristic sub-networks keep learning and training respectively and then reduce the appropriate scheme of the processing methods choosing.

ANN-Based CAPP Expert System

Process planning is in difficulty in CAPP system. It involves a lot of factors, and needs plenty of expert knowledge of which many are uncertain and ambiguous, and even conflicting. The goal of ANN based CAPP expert system is to mimic human experts' work by a computer in order to complete the process design. This imitation process applies the basic principles of artificial neural network and put all kinds of expert knowledge into a system. As a result, the system has the thinking characteristic like human and can basically design the process without the need of the participation of any human (K.Nezis & G.Vosniakos, 1997). The learning, associative memory, distributed parallel information processing function of the neural network system is used to solve problems such as knowledge expression, knowledge acquisition and parallel reasoning in the expert system.

From the point of mathematics view, the design process is expressed as: given an accessory D that to be processed. in accordance to the mapping relationship between the accessory attribute and its process output, the neural network model estimate the process route of M. Therefore, the input of the network model is the expression to show accessory design and processing attribute, and the output is the appropriate process route. And it can be shown as:

M=f(Attr(D)).

The Choosing of the Network

The basic principle that the BP network model to process information is: The input signal Xi act on the output node through the intermediate nodes (hidden) and produce the output signal Yk by a nonlinear transformation. Each sample of the network training includes the input vector X, the expected output value T and the deviation between the network output value Y and the expected output value T. By adjusting the linking intensity between the input node and the node of hidden layer,the linking intensity between the node of hidden layer and the output nodes as well as the threshold value, the error decline along the gradient direction. And after learning and training

repeatedly, the corresponding network parameters (weights and thresholds) are confirmed according to the minimum error and then the training will stop. After that, when entering an input information similar to the sample, the trained neural network is able to process the nonlinear conversion of the information and reduce the final information that has the minimum output error.

Process design requires the designers classify quickly and accurately and narrow the searching scope of the process knowledge, then put up logic reasoning by process knowledge. The associative memory capacity of the BP network is well fit for solving the problem of classification. However, logic reasoning can be carried out by the expert system. Therefore, in most of the CAPP systems, the BP network is taken as the basic model. Various sub-networks are set up to achieve the function of the CAPP system according to the needs of the CAPP system, and then all the sub-networks integrate with implement the output of the system.

Knowledge Expression, Acquisition and Learning Mechanism

Knowledge expression is a basic theoretical issue concerned by both artificial intelligence and cognitive science. It is the problem that how to pasteurize and formalize the knowledge. The process knowledge in the ANN is expressed by neural networks and rules in the CAPP system. The knowledge extraction mechanism is the bridge between internal and external knowledge expressing. The characteristics of the rule are that it solves specific problems with incomplete heuristic knowledge. Therefore it is fit for the condition that needs a lot of experience and knowledge to solve problems. This characteristic is very suitable for process design field and technology decision-making logic field. The use of rules needs us make full use of our accumulated knowledge. However, knowledge is not that simply, independently expressed as a certain rule. In fact, an enabled mechanism is constructed by using a large number of neurons through the connecting mechanism, and then information is expressed by a number of neurons. At the same time, one neuron can be involved in the storage of information, and the information is stored to a certain relationship. Therefore, the knowledge expression in the CAPP system can be described by the block adjacent weight and the threshold vector of the ANN. The ANN applies the BP algorithm to determine the weight values and the thresholds of each neuron via the re-learning mechanism of the BP network, so that it can implement the knowledge expression.

The process of knowledge acquisition is actually network samples' learning process. This process includes: converting the rules into sample vector; confirming the number of the input, the output and the hidden nodes as well as the network layers; designing the architecture of the network; providing network learning algorithm.

Process design is a work of high experience and pertinence. With the development of the production conditions and the production technology, new materials, new techniques and new equipment constantly emerge and the production experience also has changed. This obviously requires the CAPP system to fitting for these changes, which called for CAPP system with learning ability.

The information processing function of the neural network is decided by the input-output feature, network topology, the size of the linking weight and the neurons' thresholds of the network unit. When the network structure is fixed, the learning mechanism comes down to the changes of the linking weight. The ANN network described in this paper is based on the BP networks and formed mainly by applying BP algorithm. Through this re-learning mechanism, it generates a new set of weight distribution and thresholds, so that CAPP system's learning on the new process can be implemented the as well as extends the knowledge of the CAPP system.

The System Model

According to the ANN theories introduced before, CAPP system model is constructed by integrating with traditional expert systems. The core of this CAPP system is the knowledge acquisition machine, the rule base, the integrated knowledge base, the ANN learning machine and the flexible reasoning machine. The knowledge acquisition machine will transit the accessory information, resources information and processing information which comes from the CAD system into the form of knowledge expressed in the CAPP system. The ANN learning machine trains the samples through learning algorithms, and automatically learns new knowledge, new experience to get the weight value distribution and the weight matrix. And of course the knowledge base can be extended. The flexible reasoning machine educes several viable processing scheme based on the feedback information of the resource restriction, processing hours, processing cost restriction and job scheduling and monitoring.

The Advantage of the System

There are a few advantages to the CAPP expert system based on neural network:

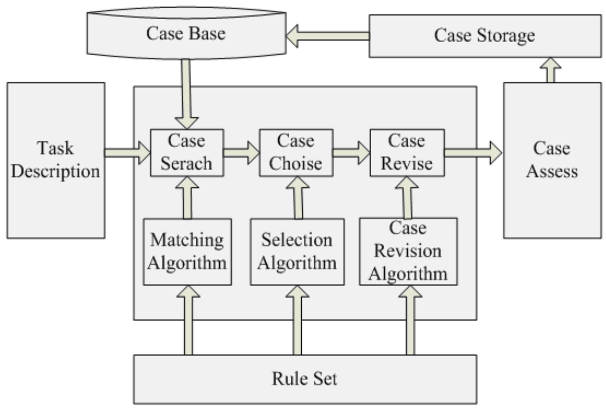
1. User does not need to organize a lot of production rules and knowledge framework, but only provide a lot of example data. And the system will acquire processing knowledge through self-organization, self-learning.
2. The reasoning process of the neural network is parallel. As a particular input mode, it can generate the output mode after internal activity. However, the output units of the output mode are all conceptualized by the logic and the output units being activated are the solutions. This kind of reasoning is closer to the thinking of the human brain. It will reduce the time that traditional expert system search for the knowledge base when it is reasoning, thus improve the speed of reasoning.
3. The input of the system is able to contain not only inaccurate information, such as the information expressed by membership degree, the probability value, the reliability degree, but also unknown information. The network will give reasonable solutions to both the two condition. It shows that this system supports fuzzy reasoning and default reasoning.

With the continuous deepening of CIMS promoting, it demands the enterprise to respond quickly to the products requirements and changes. It will also promote the study of the CAPP system. The ceaseless growing up of the artificial intelligence technology and computer technology means a lot to the intelligent, automated CAPP system. However, the real implementation of CAPP base on the actual needs of the enterprise. Improving the staff's diathesis and preparing the basic data so that the CAPP technology plays its important role in the modern manufacturing.

Compared with Traditional CAPP

The traditional artificial intelligence methods generally use the symbol logic method to construct the CAPP expert system. It analyses and summarizes the knowledge that human experts created in the field of process planning, and programs according to appropriate algorithms and rules. It is based on logical and symbol operation and implements intelligence behaviors by algorithms. This kind of design method often can not work effectively because of some problems such as matching conflict, infinite recursion and so on. However, the neural network has self-learning function. It can sum up general formulas which has no obvious rules or analytical relationship from many cases. Consequently, the network can be well applied to systems of which the relationship between input

Figure 3. The Theory of CBR



and output is fuzzy and can solve the contradictions in traditional design.

THE USE OF CASE-BASED REASONING IN CAPP

The Introduction of Case-Based Reasoning

The research of Case-Based Reasoning (CBR) originated in the book of Schank, named 'Dynamic Memory' (1982). In 1983, Kolodner implement CBR on the computer. CBR expresses the knowledge based on case. CBR uses the tips which rose from current problems facing (target case) to access the old problems (source case), and guidance the solution of the target case. CBR, a kind of Analogy Reasoning technology, belongs to inductive inference technology. It could create new knowledge, and its core idea is: Through the visit and adjustment for examples of similar problems solved in the past, reasoning to solve the target case. Similar problems have similar solutions.

The theory of CBR is shown in Figure 3:

The reasoning process is:

1. Input the demand, the initial conditions and other relevant information of the problem to be solved.
2. Search for a group of examples similar to the new problem from the case base (according to the requirements and the initial conditions of the new case).
3. Find the most similar case or a combination of multiple cases from all the similar cases. Then use the appropriate matching algorithms to get the solution to the new case.
4. If failed to transfer the cases, return to the case base and do re-case-based reasoning.

CBR has a problem-solving and self-learning ability which is a simulation of human. So in the less than 30 years from its born, it is widely used in diagnosis, geometric theorem proving, industrial design, decision-making and other fields.

The Advantage of CAPP System Based on CBR Technology

Process design often involves a large number of knowledge and solving skills in related areas, and the knowledge and skills vary when come to the experience, habits and different design conditions of designers, it is very difficult to abstract the general rule of a causal link. The result is that the

traditional rules-based process design methods success in the field of process design are rare, this is also a important reasons that get the practicality of CAPP system to be improved.

The introduction of case-based reasoning technology into the design process, constitute case-based reasoning technology design, a major method of CAPP system. And case-based reasoning of process design methods are different from the rules-based or other traditional design methods, it depends not only on areas of common knowledge, but also through the establishment of abstract relations between the problems description, under the unique experience of the past specific application design, The complete the process design task of the new cases. The work of experts in the field are more inclined to describe the experience of the process design rather than abstract the process design rules, thus it conduces to the exchange of information between technology designers and experts in the field . At the same time, the capacity of expansion is very strong in this system, with the accumulation of cases, the ability of the CBR CAPP system will continue to improve.

The characteristics in the CBR system are described as the following:

1. The search of the similar cases is the core of CBR reasoning
2. The ability of learning is a basic function of CBR system.

The new cases, which are continually revised from the CBR CAPP system, complement and extend the breadth and depth of the case base, create favorable conditions to solve problems of CAPP system in the future.

That the case-to-case process designs method is a new progress of the CBR CAPP system. There are several major advances in solving complex problems in terms of thinking: The first is the establishment of mathematical models, using numerical or analytical methods to solve the problem; Followed by the establishment of a model based on the rules or other forms of knowledge, using the reasoning method to solve the problem; Now jump over this indirect level of using the rules to abstract the knowledge model, directly using the case-to-case process design method to solve the new case. It is undoubtedly a big step forward.

The easily creating and maintaining a system is the great advantages of CBR to promote the use. The CBR system has overcome the bottleneck in the knowledge acquisition in a certain degree, the knowledge of the distortion basically does not exist, and the establishment of the system does not spend too much time, and easy to maintain. So that the process of design and decision-making had been greatly shorten.

CBR, as a relatively new field in artificial intelligence technology, offers a lot of better solution to some problems in CAPP system, and also the practical application of CBP in CAPP system should be developed and improved.

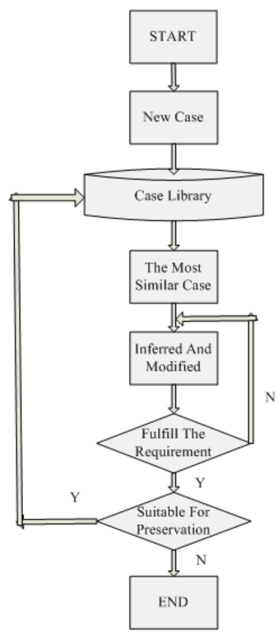
The Implement of CBR in CAPP System

The flow chart of CAPP system based on CBR is shown in Figure 4, its main ideas are as follows: First, establish an experienced case base, case base stores a number of mechanical parts and their process schemes. When new case (the new mechanical parts) arise, and new case has been compared with the case in the case base one by one, obtain the highest level similar case, that is, the most similar case. Then the process planning of the most similar case will be inferred and modified, so as to meet the requirements, the final process planning of mechanical part will be obtained. If this case is suitable for preservation, it will be deposited in case base to prepare for the future. Otherwise, not be deposited.

Establish Case Base

The establishment of a case base is precondition of CBR. This case base can store the existing

Figure 4. The flow chart of CAPP system based on CBR



cases and solutions. Quickly and accurately extracting suitable case to store case base is an important part of CAPP system based on case-based reasoning. The establishment of case base can be taken representative of the typical case to store in accordance with certain rules; Case can be composed of two following parts:

1. To CAPP system, presentation of the problem is the description of the feature information on the mechanical parts, including parts name, attributes, geometric shape, size, accuracy, materials and other content.
2. To CAPP system, presentation of solutions mainly refers to the types of mechanical parts, processing equipment, cutting tool usage and other content. The part information of encoding or short text input CAPP system through the way of human-computer interface, the use of interactive, for the calling of new problems rise.

Case Search

The powerful function of Case based reasoning system originate that it can quickly retrieve the old case which is similar to the current and new encountered problems.

Similarity is an indicator of measuring on having processing case in new parts and case base, and the similarity depend largely on how to establish the characteristics of the index, which not only with the types of parts, geometry, structure, materials and other factors, but also parts of the rough type, heat treatment method, accuracy class, geometric tolerance and roughness, and other factors. Of course, these factors of the weight should be different, the mechanical parts of the types of processing equipment, processing and cutting tool usage, and other high-impact weight should be set up in the bigger, but the weight should be set up in the smaller.

At present case index technology has major three methods: nearest neighbor, induction, knowledge-guided law or combination of these three methods.

1. Nearest neighbor method is that a user use by matching the characteristics of the right to enter the case and to retrieve the cumulative cases. This method is suitable for those who can not crawl very well-defined objectives or not available examples of problem areas. Generally speaking, only in very specific circumstances can find the same case as the new problems, more often is likely to find similar case to new problems. When a case of case base is the most similar to retrieve cases, its solutions will be proposed as the new problem.
2. Index approach is summed up in the classification of a wide range of case, summed up to determine which the best characteristics ability is. If the object of search or the results of case can be well defined and there are enough cases to summarize, so this method

is better than the nearest neighbor method. However, when the scale of using case is large, in order to complete the summary, the system needs a considerable number of cases to generate determine characteristics, summarized analysis, which will lead to long time.

3. Knowledge-based index is making use of the existing cases in the retrieval knowledge of case base to determine which features of the cases are important. As it is difficult to edit the explanatory enough knowledge, it will also be difficult to complete the comprehensive knowledge-based index in a wide range of cases input. Therefore, the use of this method combines with nearest neighbor and induction.

Case Revise

There are two main infer methods the interpretation of rule-based reasoning and inductive inference. Case-based reasoning does not depend on the rules for problem areas, but rely on past experience and examples of successful problem-solving, this is more close to the ideas and methods of human experts to solve the problem, and thus it is more easily been understood and applied. The case based reasoning for CAPP is to compare and match characteristics information of the new processing to information of retrieval part feature information, so as to decide which parts of the process design can be retained and which should be revised, which should be deleted. After completing inference, it should also be on the process design of further revision and improvement of the final solution.

Case Preserve

In case-based reasoning for CAPP system, if not determine preserved, and new examples are constantly added to the case, it will make case base more and more huge, time-consuming retrieval process seriously. Therefore, in order to facilitate the retrieval and management of the case, for example, as far as possible to include both parts of the typical characteristics of the design and decision-making process, to avoid similar, repeated examples of storage led to the example of the Rongza. Case based reasoning has three following storage methods:

1. Associated storage in this way, all the characteristics of the cases will be indexed.
2. Levels of storage features case of this approach to organize from the general to the specific structure of the concept.
3. Determine the method of networking between these two methods, system requirements for flexibility, the use of associated storage; retrieval tasks very well defined, in order to save time for retrieval, storage levels can be organized way.

MULTI-AGENT-BASED CAPP

For a long time, CAPP has always been the bottleneck and conditionality in manufacturing system autoimmunization, and there is still some distance between most systems already developed and methods mentioned and factory practice. In modern integration manufacturing system, CAPP should include:

1. good flexibility, extensibility, reusability, distributed collaboration, dynamic adaptability.
2. CAPP system can work together with CAD, PPC (production planning and control) to make technique plan to adapt different specific working state, and realize general optimization of product design, manufacturing and production.

The practical research of Multi-Agent system began around mid-1980s, and there is an obvious

increase in recent years, its coverage contains manufacturing, industry processing control, air traffic regulation, noise control and so on. So far, Multi-Agent and Intelligent Agent have been the research hotspot worldwide. The introduction of Multi-Agent system into CAPP system can largely improve traditional CAPP system structure, enhance CAPP system performance, ensure that CAPP system can play a larger role in technique design..

CAPP Agent Modeling

In CAPP system, the inputting geometric information and technique information are produced by CAD system, CAPP system based on Multi-Agent will produce technique and input its result into CAM system according to geometric information and technique information of product model and by collaboration among Multi-agents.

MAS(Multi Agent System) is made up of multi-agents, each of which has its own knowledge library and inference engine, all Agents communicate and collaborate with each other according to well-defined protocol in order to cope with complicated problems together. That way, the whole knowledge resource in the system is made full use, avoiding singleness of single expert system knowledge library, and it's beneficial to deal with complicated problems which concern reasoning problems in many fields and multi-level, meanwhile, the operating efficiency and parallelism of system can be largely enhanced by using distribution of reasoning.

In order to satisfy the need of modern integrated manufacturing system, an intelligent CAPP system structure based on MAS is put forward, as diagram shows .All Agents collaborate with each other to make up a complete technique plan. According to the requirements of CAPP system in practical environment, each Agent can be distributed on the same computer or on different computers.

The division principles of Agent in CAPP system are as follows.

From the perspective of software engineering, Agent can be considered as the natural extension of object-oriented programming. Object not only provides data packaging, but also modular modeling and it can construct system according to mutual effect between entities. On the other hand, object is passive entity, it doesn't change its own state until its response is activated (for instance, call member function), it is not controlled before being activated .On the contrary; Agent is active and can activate itself. Each Agent has its own thread, independent of other Agent. Different Agents interact with each other by sending and receiving messages, but these messages are not done by calling functions. Messages can be broadcasted to a group of Agents, each Agent can decide independently whether it ignores or deals with this message in a latter time. With the concept of Agent, a complicated system can be divided into several different effective Agents, which are more easily dealt with. In this way, the flexibility, extensibility and stability of CAPP system can be well enhanced.

The performance of traditional technique design system is realized through structured programming thought. According to Agent thought, in technique design system the technique planning and decision are realized by constructing effective multi-level or multi-related Agents. Thus, the key technique of Agent technology in technique design system practice lies in constructing different function-independent Agents, that is, the decomposition of technique design Agent, the followings can be used as decomposition:

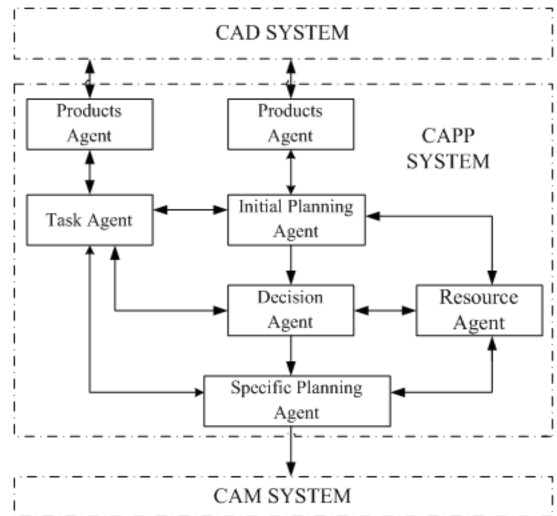
1. **Decomposition based on task.** Many behaviors in CAPP system are driven by task. For instance, to get product order from market, to determine product design task, to complete product component list, then to design technique project and technique equipment, and finish corresponding files. and so on.

2. **Decomposition based on resources.** Manufacturing resources are one of the most concerned questions in technique design process. The dynamic characteristic of manufacturing resources influence the decision of technique project, it is also one of major reasons that result in variety of technique project and complication of technique decision.. Manufacturing resources contains equipments, knives, fixture, measuring tools and so on. To consider manufacturing resource as the first Agent, manufacturing resources is divided into many next-level Agents.
3. **Decomposition based on object.** Considering that the object of technique design system is to produce goods, so product design is considered as the first Agent. Furthermore, product design Agent can be decomposed into product characteristic Agent and material Agent and so on.

The division principle is mainly based on functions and size. Module with certain function will be divided into corresponding Agent. however, the Agent shouldn't be either too large or too small. If too large, it's hard to configuration and hard to modify. If too small, more interfaces have to be defined so it's difficult to integrate. In a word, to decompose different level of Agent according to the function, task and object of technique design system. But, the refinement degree of Agent decomposition has to be taken into consideration. Over-refinement will result in low operating efficiency of the whole system. In practical Agent decomposition, proper size and scale have to be made sure of.

In this part, the CAPP system based on Multi-Agent can be divided into seven main Agents, as shown in Figure 5. Such as task Agent, initial planning Agent, decision Agent, detailed planning Agent. These Agents can be mapped into different functions and objects in technique planning, it can be organized dynamically, and to finish tasks in a flexible and collaborative way.

Figure 5. The module structure diagram of CAPP based on multi-agent



The module structure diagram of CAPP based on Multi-Agent is shown in Figure 5.

Besides common characteristics such as distribution, openness and autonomy, the distributed technique planning system based on Multi-Agent also contains following characteristics:

1. **Reconfiguration.** Because of modular structure, it's easy to reconfiguration to adapt new manufacturing environment.
2. **Customization.** It contains the function of customization. Customers can interact with system in any layer of the whole three layers.
3. **Admixture.** Its structure is made up of equal relationship (such as task Agent and initial planning Agent) and layer relationship (such as planning Agent).

The main functions of different Agents in technique design system based on Multi-Agent are as follows.

Products Agent provides geometric information and technique information of product for technique decision system, including different product information for technique design. C is the management coordination center of the system .It is an open and systematic coordination environment, ensuring that customer can interact with system to evaluate the product design at the beginning of product design, meanwhile, to coordinate selectively requirements of beginning of product design and feedback to CAD system. Task Agent has the ability to recognize different information, like human being's geometric reasoning ability, it is responsible for changing design data to specific language information for manufacturing, and accepting product model and changing it to the public model which is shared in the inside of CAPP system. According to the product project information provided by product Agent, Task Agent decompose technique task and send tasks to corresponding Agent to operate, and control decision order and process of CAPP, and respectively trigger initial planning Agent, decision Agent and specific planning Agent to work according to three different processes.

Technique planning Agent is composed of three Agents, that is, initial planning Agent, decision Agent and specific planning Agent, each of which has its own knowledge for specific technique planning task in three different processes.

1. The input data of initial planning Agent includes two types: one is product model from design system, to show geometric shape, design tolerance and surface roughness and material characteristic. Resource information from technique design system. The task of initial planning Agent is to make use of component information and resource information to search all practical operation. This means that initial planning Agent doesn't finish all the technique planning, it only produces many operation chains based on characteristic. The output of initial planning Agent is a set of operation chains .According to workshop resource information from production operation feedback, to allocate them to machine tool. Actually, all the operation chains available make up technique way available. Another task of initial planning Agent is to work as a tool of designer to evaluate design project, designers can check possibility of production, modify their design project according to workshop's environment, so that the possibility of production and profit is enhanced. The difference of possibility production evaluation of technique design system from other ways lies in that it not only considers design (that is technique ability, it can be described through component shape, size, tolerance, surface roughness, geometry and technique condition and technique profit)but also considers all the conditions of production environment, to analyze the possibility of production according to resource information in the workshop.
2. Decision Agent is the center of whole distributed technique design system, as diagram 2 shows. As for traditional CAPP system, people made specific design just shortly after choosing a theoretical optimized technique project according to technique, cost and so on . Because of not considering the production and management from the perspective of whole operation processes of production workshop, It is quite easy to cause bottleneck of some resources in the workshop and people have to renew the technique design .The main task of decision Agent is to choose proper technique project for workshop level according to dynamic information of workshop and up-level technique project. Generally speaking, there are many projects available for component operation and there are many ways for characteristic operation, In order to make sure of practicability of technique planning system in the workshop

level, when making technique system decision, shipping date of components and practical operation conditions (for instance, resource availability) in the workshop have to be considered. To choose operation methods and resources properly is of great value for enhancing operation flexibility and keeping operation going on smoothly. For instance, when shipping date is at hand, high-quality operation machine tool have to be choose to make sure of shipping before the deadline. On the contrary, some economic ways can be choose. In this decision level, according to resource information from technique analysis and real-time information sheet of machine tool data(provided by resource Agent),to make proper resource match in accordance with priority of technique project.

3. Specific planning Agent. Specific planning is done before production, according to optimized object, make sure of shipping date and resource available, to do specific technique design, check cutting parameter, operation time and NC program and so on, output complete technique plan for guiding workshop's production.

Multi-Agent Coordination in CAPP

As mentioned previously, the CAPP system environment is organized as a community of cooperative problem-solving agents, where each P-agent is a relatively independent and autonomous knowledge- based expert system. The P-agent solves problems in its limited domain independently. Therefore, it should have the capabilities to act as a member of a community. These capabilities include:

1. A shared communication language with other agents.
2. Internal knowledge representations which capture sufficient goal and history information to allow.

For solution revision to be carried out cooperatively, it includes:

1. Provisions for sharing information in a timely manner for problem-solving.
2. Mechanisms for incorporating externally produced partial solutions.
3. Mechanisms for negotiation to settle conflicts.
4. The ability to coordinate an internal agenda with external events.

Above paragraph introduces how agent coordination requires agent itself to be, the coordination principle between agents is as follows:

While each Agent in the CAPP system is working, coordination requirement have to be put forward according to technique design task, and set up coordination object, and take charge of the operation, planning and division of tasks in the coordination body, and communicate with other intelligence entity, meanwhile, to choose suitable coordination partner in accordance with competitors' conditions in the network .During the coordination, each firm or factory takes its profit as the biggest object, and choose comparatively lower cost member (competition winner) as partner, so as to form coordination relationship.

The main process of son-task division way based on contract net: first, coordination starter broadcast son-task to all active Agent, then each Agent decide by itself whether to bid or not according to its own ability, interest and present load. During an effective period, coordination starter uses arbitrating algorithm to choose suitable partner among so many bidding documents, and then sign contract and divide son-task. The advantage of contract net lies in simple system structure and good extensibility, but low efficiency, because it takes up a lot of broadcasting resources. In order to cope with its disadvantage, an acquaintance model in the CAPP system is designed to deal with task division. It can be defined as F=<I, F, S>, I stand for Agent identifier. F is aptitude tendency. S is

the complex evaluation of Agent history. During interactive communication between Agents, a acquaintance list is set up, when bidding, first to take son-task requirement as searching condition, coordination starter looks for the Agent which has that aptitude tendency in the acquaintance list, and make drop-down list according to S value, then, ask all Agents one by one from the largest number if they like to cooperate again. After the bidder finishes the task, the complex evaluation of Agent history is adjusted according to the quality of the finished task.

The two methods mentioned above are not contradictory, instead, they are complementary. When running the system the first time, all the initial values are zero, broadcasting of the contract net is used to find partner, after bidder's finishing the task, information of the Agent is added the into acquaintance list. That way, a period of accumulation later, when dividing task, acquaintance priority strategy is choose to enhance the whole communication efficiency of the system.

The introduction of Agent into CAPP field is to make use of distribution, autonomy, coordination and activeness of Multi-Agent to find the solution of problems, to discuss the intelligence problem. In the technique planning system based on Multi-Agent, according to functions or objects, the technique decision system is divided into several single Agent intelligent identity with independent function, each of which is independent to each other as well as related with each other .Since different Agent has different description of problem and different solution, and because of different problem-solving perspective and evaluation of the same object in different Agent, the conflict is unavoidable

According to technique design system structure model based on Multi-Agent, the conflict of Agents can be divided into outside conflict and inside conflict from the perspective of where the conflict is caused. Inside conflict is mutual contradiction and inconsistence of Agent inside different knowledge, planning or condition such as the same characteristic surface as well as decision rule repetition, intercross and contradiction. Outside conflict is the mutual relationship of independence as well as restriction among Agents in the process of technique design. When one Agent is solving the problem, there is no doubt that It will destroy some restriction, so the conflict happens. The conflict solution of Agents is also an important part of Agent coordination; meanwhile, it is one have to be taken into consideration.

In practice, a lot of conflict solutions show in the form of rules. To use making rule to show conflict resolution rule. Conflict solution rule is composed of precondition and conflict solution methods. That is:

If (precondition) than (conflict solution method)

Precondition is description of conflict, which explains the type of conflict and object; Conflict solution is to give solution advice according to precondition, which is conclusion of experienced experts in the related field and often taken by them in their work. The followings are rule examples for technique conflict solution:

- **Rule 1** If cost exceeds budget and operation method can be modified, Then use other operation method to replace.
- **Rule 2** If the operation method of some characteristic is broaching; then search manufacturing resource, choose machine tool
- **Rule 3** If the operation method of some characteristic is broaching and there is broacher in manufacturing resource, then choose specification of machine tool.
- **Rule 4** If the operation method of some characteristic is broaching and there is broacher available in manufacturing resource, then change operation method.

The Characteristics of Agent-Based CAPP

The traditional CAPP system and design methods have shown much weakness, such as poor adaptability and openness, long development cycle. Thus, to research and develop an open structure-flexible CAPP system has become today's hotspot. Agent is an entity that includes characteristic of activeness and inter-activeness. MAS system, which is made up of Agents, contains the characteristic of distributed parallel processing robustness and openness and it can make fast response when emergency breaks out, meanwhile, it can adapt to changing environment. In nature, it is distributed and extensible. Numbers of Agent can be added if it is necessary in order to enhance system function. Multi-Agent is easy to maintain and modify and extend. Thus, Multi-Agent structure can be used to set up CAPP system

CASE STUDY

Facing the increasingly fierce market competition, we have mentioned the importance in the above sections that the enterprises should improve the rapid response to the process as well as develop and implement the CAPP system. At present there are a few commercial CAPP system, however, these CAPP software have common characteristics:

1. Just suitable for mass production of mechanical processing enterprises.
2. Focus on process design, be weak in production management.
3. The CAPP software is too expensive whose features are perfect, and intelligent level is high. For the above characteristics, these CAPP systems do not meet the needs of the small workshops. At the same time, they don't have better management functions either.

In this section, we will give a small CAPP system based on the production management.

In this system, we regard the parts production flow as the main line according to the practical needs of the workshop producing, integrate all the information needed in the producing process in the workshop so as to implement the sharing and timely updating of the information. By doing this, we can improve the quality of the process management, process management and production management and finally solve the problems such as the process design, management, query, changing, and tracking completely. Also, we can reduce the repeating work in the whole production flow from the drawings to the products, improve the efficiency of technical management and production management in the production flow, and finally improve the productivity.

System Framework

CAPP system is a complex integrated system which includes techniques such as form processing, word processing, graphics processing, database accessing and so on. Therefore, its main function is to ensure the integrity and consistency of the products process data, implement the integration and sharing of the products process information, rather than design the parts process flow and output the process card all alone. In this section, we bring forward the functions that the system needs to achieve basing on the analysis of the production fact in the small workshop. By applying the advanced network and database technology as well as taking into account the contact among all the functional modules, we can design a small web- and- database- based CAPP system. It uses the database technology to design the process database and process resources database in order to implement the effective management of the process data and resources and improve the quality and efficiency of the process design. By using the Internet platform, we can manage the entire production process for the workshop to, such

as the effective management of process, the sharing and integration of the process resources, the convenient statistic and gathering of the data.

The system establishes four sub-systems: the process management system, the statistical management system, the completion inspection system and the materials information system.

1. **The process management system:** Mainly used by the technologists of the workshop. According to the authority size, the technologists of the workshop can search and consult the standard process documents as well as build the process flow of their own need after a series of operations.
2. **The statistical management system:** Used by the statisticians of the workshop. The statisticians are able to enter the staff information library to view and modify employee information, process the statistics of the man-hour and the plans' completion situation, search the man-hour and the plans' completion situation for a certain part. All this function will supply data for the workshop management.
3. **The completion inspection system:** Used for the checkers of the workshop. The checkers judge the results' eligibility of each working procedure and then check in.
4. **The materials information system:** Used by the material controllers to manage the materials of the workshop and the storage. The material controllers must register the materials used during the production and analyze all the materials storage of the workshop. For example, when some materials are lack of inventory, they need to make the purchase application and after the materials have purchased, they should update the materials library in time.

Introduction of System Software Structure

C/S (Client/Server) structure, that is, client and server structure. It can make full use of the advantages of both the two parts' hardware environment and distribute the tasks rationally to the Client and Server so as to reduce the communication costs of the system. In this structure, the client completes the expression logic of user interface and the business logic of the application. However, the data services (such as the addition and deletion of the data, and some other operations) are completed by the database server.

Introduction of the System's Database Server

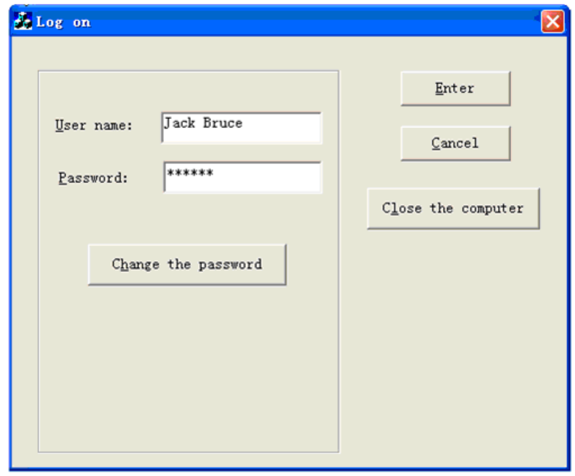
For no matter which kind of system, we need not only the friendly man -machine interface running on the client, but also a good database system running on the server. Well-designed database will run faster, operate and maintain the data more easily. MySQL, designed for the speed and the stability, is a relational database management system (RDBMS) of high-performance, multi-threading and multi-user. The advantages of MySQL database include: superior performance, reliability, ease of use, support most important features of the ANSISQL- 99 standard and support multi-user.

The system established data tables such as the staff' basic information table, the materials inventory table, the equipment table, the tools table, the process and equipment table, the process rules table. In addition, the system has given the relationship between these tables. In structure, we have blended all the information of the production management together to form a database of the management information system (MIS).

In the following parts, we will briefly introduce the function and the structure of the basic tables in our MSI database:

1. **work type: work type table.** Record all the work types of the workshop, including: management, technology, lathing, milling, planning, numerical control lathing, numerical control milling, line cutting, EDM, welding, heat treatment, machine repair, lining, clamping, electric repair and so on.
2. **my group:team information table.** Record all the teams of the workshop, including: the management team, the technology team, the NC team, the lathe man team, the milling and planning team, the locksmith team, the welder team, the heat treatment team, the machine repair team, the electric repair team of 10 teams. The structure of the table is shown in the following table:
3. **Standard-art-information: typical components table.** Record the main containing information of a typical component, such as the index number of the work piece, the order of the part, the model of the part, the name of the part, the graphic No. of the part, the unit, the quantity, the name of the materials, the specification of the materials, the department commissioned to process, the issue date, the planning completion time, the actual completion time, the compilation, the proofreading, the auditing, the approval, the graphics of the part, the graphics of the process and so on.
4. **Status handlers: the statistics of the completion information.** Record the process's completion status of each ordered production's parts, such as the index of the working procedure, the index of the completion records, the operator's name, the completion hours, the inspection date, the name of the inspector, the quality status and so on. The information of the man-hour management is picked up from the table status-handlers, and the three tables contact with each other by the work piece index, the process index, and the completion records index.

Figure 6. Log-on interfaces



Application Development

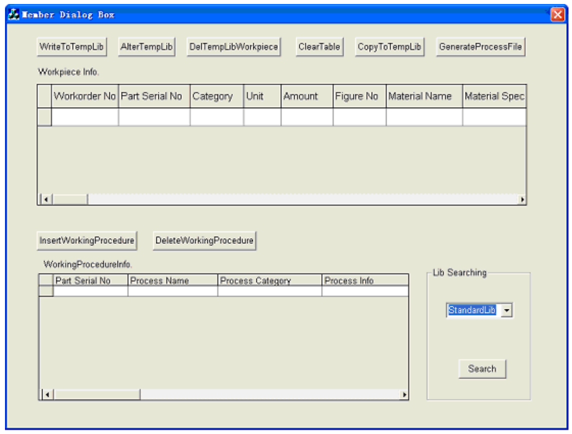
The Design of the CAPP Security System

We developed the programmer diagram of the CAPP security system as shown in Figure 6.When a user enter the system, the system will verify the user's permission and the password. The user can enter the system only when he has input the right user name and password. If the user is lack of the permission or user has filled in the wrong user name and password, the system will give an error message in order to require the user to fill in again.

The Design of the Process Management System

The process management system is mainly used by the technologists of the workshop. When the system is going to design the process flow, it will simulate the habits and courses that the technologists design the process. Firstly, it will find the process flow of similar parts in the standard process library to see whether there is something used for reference. Secondly, determine the process route, modify or re-design the content of the process,

Figure 7. The member dialog box



and draw the process diagram. And finally print out the process flow. It provides functions such as: search typical process from the standard process library or temporary library; design the process flow; insert, delete and clear your processes; draw the process diagrams; generate and output the process documents.

The member dialog box of the process management system is shown in Figure 7.

The Design of the Statistical Management System

In the actual production, in order to understand and master the production in time, we need to gather and sum up all kinds of information, such as working hours of workers, materials stocks and so on. The statistics of this information are completed manually before; therefore, it is inefficient and prone to error. The function of the statistical management system is to extricate the statisticians from the tedious statistics tasks as well as ensure the speediness and accuracy of all the statistical tasks. The statistical management system designed three modules: searching and altering function for the statisticians, recording man-hour function and recording staff information. The statisticians search and alter the major statistical information of the work piece. Meanwhile, in order to query the process of a part or the detailed completion information, when checking a work piece, it will automatically add a row in the man-hour information table and display the process and the completion information of the work piece. In the program, the statistics of the work piece and the statistics of the man-hour are separate. By using the employee information module, we can complete the operations such as adding, altering and deleting to the employee information.

According to the actual needs of the workshop, we designed more than 20 functions about man-hour statistics. Although these functions are different, their programming idea is basically the same. We can generally achieve them in three steps:

a. According to the different statistics functions, write different SQL query sentences.
b. Offer the query function the query sentences, and query the results from the database.
c. Display the query results in the table.

As MYSQL is not support for the in-table searching currently, we cannot get the query results after searching the statistics only once. And only can we turn to the VC program function for managing and preserving the during the query operations. And then we query again according to the intermediate results. After a few times of

querying and calculation, we can finally get the data needed.

The Design of the Completion Inspecting System

As previously stated, in order to ensure the effective operation of the whole management network in the workshop, make sure that the leaders and dispatchers of the workshop can monitor and understand the progress of all the products, inspect the implementation of the plan and adjust it constantly according to the actual situation and needs so that we can achieve the aim of completing the task, the checkers must register the eligibility of the products and the man-hour completion situation, and then store the completion information in the database.

In the design of the completion inspection system, we mainly consider about the following aspects.

1. We must ensure the security of the data. The checkers cannot add, alter and delete the information of the work piece or the working procedure. A checker cannot alter or delete the completion information that another checker has preserved. At the same time, in order to prevent misuse, proper tips should be given when somebody is altering or deleting the completion information. The operator's name being input should be correct and the completion working hours are not allowed to exceed the planning working hours. Once there are some mistake, tips should be given before the data or information are written into the database.
2. Since the workload of the verifying is large as well as some checkers are weak in the input of Chinese characters, we consider about that the checkers did not enter Chinese characters to complete the operation in the design. We give all the names of the teams and staff in the user interface so that we can greatly improve the convenience and speediness of the inspecting operation.

SUMMARY

Along with the increasingly fierce manufacturing industry competition, emergence of new craft and technique, diversification of product requirement, reduction of product's life cycle,the development and application of CAPP system is displaying increasing importance and urgency. This chapter introduce the development of CAPP at first, then on the basis of CAPP theory, introduces and analyses several algorithms and technologies, which have been applied to CAPP, such as genetic algorithms, artificial neural network,, and technology of agent.

REFERENCES

Abid, S., Fnaiech, F., & Najim, M. (2000). Evaluation of the feed forward neural network covariance matrix error. In *Proc.IEEE ICASP 2000*, June 5-9, Istanbul,Turkey.

Chang, T. C., Richard, A. & Wysk, Wang H. P. (1998). *Computer Aided Manufacturing*. New Delhi: Prentice Hall of India.

Devedzic, V. (1996). Organization and management of knowledge bases—An object-oriented approach. In *Proc. Third World Congr. Expert Systems*, Seoul . *Korea & World Affairs*, *II*, 1263–1270.

Devedzic. (1996). Organization and management of knowledge bases-An object-oriented approach. In *Proc. Third World Congr. Expert Systems*, Seoul, Korea.

Durkin, J. (1994). *Expert Systems-Design and Development*. New York: Macmillan.

Mou, J. (1997). A Method of using neural networks and inverse kinematics for machine tools error estimation and correction. *Journal of Manufacturing Science and Engineering*, 119.

Nassehia, A., Newmanb, S. T., & Allen, R. D. (2006). The application of multi-agent systems for STEP-NC computer aided process planning of prismatic components. *International Journal of Machine Tools & Manufacture*, *46*, 559–574. doi:10.1016/j.ijmachtools.2005.06.005

Nezis, K., & Vosniakos, G. (1997). Recognizing shape features using a neural network and heuristics. *Computer Aided Design*, *29*(7), 523–539. doi:10.1016/S0010-4485(97)00003-1

Rocha, J., Ramos, C., & Vale, Z. (1999). Process planning using a genetic algorithm approach. *IEEE International Symposium on Assembly and Task Planning 1999*, (pp. 338-343).

Sparrow, F. T., & Bowen, B. H. (2005). The Future of SAPP, WAPP, CAPP, and EAPP. In IEEE PES GM 2005 Panel Session, San Francisco, June 12-16.

Xu, P., Gao, Q. & Wang, M.X. (2006). GA-based CAPP optimization strategy. *Chinese journal of construction machinery, 4*(3), 303-306.

Yi, H., & Ni, Z. H. (2003). Study of CAPP Technology in Networked Manufacturing. *Computer Integrated Manufacturing System, Beijing*, *9*(3), 96–99.

Yip-Hoi, D., & Dutta, D. (1996). A genetic algorithm application for sequencing operations in process planning for parallel machining. *IIE Transactions*, *28*, 55–68. doi:10.1080/07408179608966252

Zhang, F., Zhang, Y. F., & Nee, A. Y. C. (1997). Using genetic algorithms in process planning for job shop machining. *IEEE Transactions on Evolutionary Computation*, *1*(4), 278–289. doi:10.1109/4235.687888

Zhang, J.G. & Li, P. (2003). G.A holonic architecture of the concurrent integrated process planning system. *Journal of Materials Processing Technology, 139*(1~2), 267-272.

Chapter 12
Intelligent Diagnosis and Maintenance

Nowadays, intelligent diagnosis for the complex system has been a forefront of issue. The application of the artificial-intelligent technology has made the dream of using men's knowledge to diagnose the complex system, and improves it up to a new grade.

Traditional diagnosis is up to technical engineer experience to estimate equipment's status, and make a judgment; this way has many limitations and little efficiency, with the increase of complexness of the equipment, there have to get some more effective methods.

In 1980s, some experts have researched on the diagnosis system using intelligent technology. With the development of computer and network technology, intelligent technology has better support platform. Experts have researched on different branches of diagnosis technology and used these ideas into diagnosis system (Wu J.P. & Xiao J.G., 1997).

As the developing of Internet, all equipments have been cyber and connected to network. More and more system is consists of multiple devices and have the characteristic of the distribution. In this case, remote diagnosis system gets more attentions because of its unmatched benefits. In the remote diagnosis system, the technology of multiple agent or called multi-agent often be used to resolve difficult diagnosis problems.

Intelligent diagnosis is the develop trend that could perform intelligent maintenance in a high level of efficiency. Researching on intelligent diagnosis and applying it have significant meaning.

Definition and development of diagnosis is introduced in this chapter firstly, which including different branch and fussing with other area. The advantage and shortage of different technology also be introduced. Then the remote diagnosis on network is discussed. The theory and development of Multi-agent based remote diagnosis technology is also presented. The trend and scene is been bring on finally.

INTRODUCTION

The fault diagnosis of equipments is from the military needs, with the development of micro-

DOI: 10.4018/978-1-60566-864-2.ch012

electronics technology, computer technology and sensory technology, mechanical equipment fault diagnosis technology has been more perfect and develops towards intelligence. Signal analysis of the various means had provided a good opportunity for the development of fault diagnosis technology (Wu J.P. & Xiao J.G., 1997).

The mode of equipment monitoring and diagnosis has a development process which is from single monitoring and diagnosis systems to distributed monitoring and diagnosis system (DMDS), and then to Web-based remote monitoring and diagnosis system. In recent years, the southern United States power companies, IN-LAND Iron and Steel Company, QUANTUM chemical companies, BENTLY companies have all developed their own network status monitoring and diagnosis (Zhou Z.D., 2004).

Definition of Fault Diagnosis

Fault diagnosis is that the system identifies some causes and features which lead to a functional fault in a certain circumstances, and make a judgment on where the fault state occurs. The basic idea is expressed in general like this: assume that the detected objects all status that may occur consist of a status space X. and its measurable features range consist of features space Y(Zhou Z.D., 2004).

When the system is in a status x, the system has a definite feature y. There is a existence of mapping: g: $x \rightarrow y$. Instead, certain features also determine the definite status, namely, the existence of maps f: $y \rightarrow x$. The purpose of fault diagnosis is to determine what the condition is basing on measurable characteristics of vectors of the system.

Fault diagnosis theory is a emerging discipline that is based on reliability theory, information theory, cybernetics, control theory, system theory. It used the modern testing equipment and computer technology as means, digesting various object special laws for its theory (Zhang Q.B.,1997). There is three parts, the research on Physical and chemical processes in fault diagnosis, the research on fault information discipline which is concerned with the collection, selection, processing and analysis of the fault signals, the research on the logic of diagnosis and mathematical principles, mainly uses logic methods, to figure next detecting part.

Along with manufacturing system's development, the monitoring and diagnostic and maintenance technology is the main access to get information in the whole life cycle of manufacturing system, and its role is to ensure that the manufacturing operation of the system is safe, reliable and technical acceptable and make it meeting economic requirements.

Definition of Intelligent Diagnostic System

Intelligent diagnostic system is defined as a system which consists of man, device and software. It is aim to diagnosing identifying and predicting status of the diagnosed objective.

1. It's an opening system and its capability is progressed in the process of application and interaction with environment.
2. It is a system consists of computer hardware and software. Unlike the traditional computer system, it has indefinite algorithm and procedure approach. According to requirement of diagnosis process, intelligent diagnosis system search and utilize the knowledge and experience of expert to diagnose.
3. It is an artificial intelligent system includes some hardware and software meanwhile is not limited in the traditional computer which has the limitation of local information storing and serial symbol dealing. With the development of ANN, diagnosis system adapt it and use the learning function of ANN, associating memory, distributed parallel information processing to solve the prob-

lem of knowledge presentation, knowledge acquisition and parallel reasoning.

Intelligent diagnosis technology tries to simulate the human being expert to diagnose the complex system. Its advantage is integrating experiment of many experts, so its capability is exceed expert or at least same with experts.

Development Stages of the Fault Diagnosis

According to diagnostic tools and methods, mechanical fault diagnosis technology development has experienced three stages:

1. Firstly, the initial stage when the result was determined by the sensory experience and professional knowledge of the experts in the field.
2. Secondly, the modern diagnostic stage when the diagnosis process is based on computer-based signal processing, and using sensors and dynamic monitoring technology as a mean (Wu J.P. & Xiao J.G., 1997).
3. Thirdly, integrated and intelligent diagnosis stage when the system has integrated monitoring function, diagnosis function, management function, and scheduling function.

According to the range of using, its development has three stages.

1. **First stage is the separation unit stage.** Early manufacturing was not in a high degree of automation, neither had yet formed systematic integrated, so the diagnosis was mainly used for single device or single type. Monitoring system and diagnosis system exist as separated modules, and they just performed single or few function like only monitoring or only diagnosing.
2. **Second stage is the combined unit's stage.** With the increasing of the complexity of system, the corresponding demands for monitoring and diagnosis for manufacturing process and manufacturing device has been a hot topic. Therefore the monitoring unit and diagnosis unit had been a part of the manufacturing system.
3. **Third stage is the stage called integrated system.** A new generation of manufacturing systems which are represented the FMS, CRMS, and other advanced manufacturing systems which are large scale, more functional, and more complex structure. They need an integrated monitoring and maintenance system to diagnose, control and manage. This results a distributed monitoring and diagnosis system which ties the distribute different parts of the sensors and a wide range sites of services and management, and forms a computer-based LAN(Local Area Network) (Zhou Z.D., 2004).

Those systems mentioned above are local-oriented devices or systems. Its monitoring, diagnosis and maintenance are accomplished locally. Therefore they all have regional characteristics. Once a fault occurs but it could not be resolved by local diagnosis system, the factory has to power down and shut the manufacture process down to look for helps from the provider. This spend a lot of time and financial resources and the capability of rapid reaction and the effective working hours have reduced greatly because of the long time lay-off of the devices or the systems. In this case, the diagnosis and maintenance of device and system is inconvenient.

New Requirements to Modern Diagnostic System

Modern manufacturing is oriented to the general direction of development of flexible integration, intelligent, distributed, reorganized and the global manufacturing (Qian Z.Q.,2004). And it concerns the access to the collection, transmission, proces-

sion, usage of the information, and other works as the core, continuously promote towards the breadth and depth. In the new manufacturing mode, different manufactures in different regions need to establish dynamic alliance, compose flexible virtual corporation that implements co-design and remote manufacturing. These virtual enterprises have a notable feature that they could response to customer's demand rapidly, and they could shorten the time of products to the market. In such circumstances, normal working is the key factor, so the monitoring and diagnosis and maintenance for the system is important (Li M.& Wei Z.S.,2007).

Monitoring and diagnosis system have the following trends.

In the architecture, the MDS must have a distribute and network architecture to meet the need of agile manufacture(Li M.& Wei Z.S.,2007).

In the functional structure, inspection and monitoring system in various modules have to be relatively independent and has a good open and self-discipline in order to ensure manufacturing system distributed architecture and high flexibility, the modules that could be any combination of both can be integrated with other systems, but also complete its independent function(Qian Z.Q.,2004).

In the physical structure, inspection and monitoring system should be modular and "plug-in" structure and could always be reorganized through the network and software in order to meet the manufacturing system of globalization, virtual and organizational structure of the reorganization, and other requirements.

In performance, since inspection and monitoring system is the sensory organ of advanced manufacturing systems, the work of the inspection and monitoring system's reliability, the authenticity of the information are both very important. The robustness of system includes fault tolerance, online maintenance, and intelligence faults quick diagnosis and so on.

Internet-based remote monitoring, diagnostic and maintenance system is a product of above trend. It resolves the remote equipment monitoring and fault diagnosis, and ensure that the remote complicated equipment or system being normal and efficient performing through the exchange of information rather than the exchange of men. Because of its practicability, fastness, convenience, and other advantages, it is called as a new generation of equipment condition monitoring and fault diagnosis system, and also a new field of research(Li M.& Wei Z.S.,2007).

DIAGNOSIS TECHNIQUES

Many methods of intelligent diagnosis discipline are closely linked to the theories of artificial intelligence technology, such as the rules-based, knowledge-based, artificial neural network based, case-based, behavior-based, machine learning methods, and fuzzy logic and such on. Although these methods have different features and advantages, there are many limitations of their own (Wu J.P. & Xiao J.G., 1997).

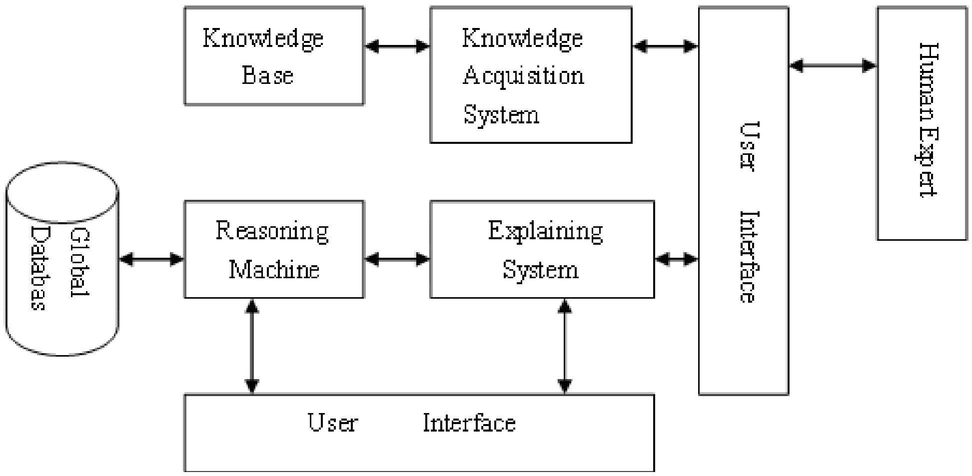
Many branches are there in the field of artificial intelligence research (Wu J.P. & Xiao J.G., 1997) There are two major branches used in the field of equipment diagnosis, namely the knowledge-based diagnostic expert system and neural network based intelligent diagnosis system.

The superiority of intelligent diagnosis lies in its integration of the best experiences of a number of experts, these system's function level can be achieved at least the level of experts or over the level of experts, and they can achieve rapid analysis and diagnosis for multi-faults, multi-course, and sudden fault.

Knowledge-Based System

Knowledge-based intelligent diagnosis technology is the most mature direction of development in the field of equipment diagnosis, and is also

Figure 1. Structure of expert system



the class of intelligent diagnostic technology which has the more research attention and the most extensive application (Wu J.P. & Xiao J.G., 1997). It generally has experienced two stages of development: the first-generation of fault diagnosis expert system based on the shallow knowledge (knowledge of human expert's experience), and the second generation of fault diagnosis expert system based on deep knowledge (object model of knowledge) of fault diagnosis expert system. And the mixed structure of expert system, is a combination of these two methods, complementing each other.

Shallow knowledge-based fault diagnosis system is at the core of the enlightening experience of knowledge of experts and operators in the field, it get the diagnostic results through deductive reasoning or creative reasoning, with the aim of finding a set of fault that it can give the best explaining for a given collection of symptoms (Hamscher, W., Console, L., & de Kleer, J,1992). Generally in its reasoning process, there is two types of knowledge must be used, one type is the knowledge which interprets the relation between fault and symbol of status, other one type is the numerical knowledge which reflects the degree of causal relationship between fault and status (Chr. Ange li.,2001). Over the years, people have established many reasoning methods, with a combination of these two types.

Expert System summed up the past experience of experts in the field into rules, and used it to reason in the purpose of fault diagnosis. It has these parts like knowledge base, reasoning machine, knowledge acquisition subsystem, user interface, explaining mechanisms, and global database. The expert system structure is shown in Figure 1.

The expert system structure includes below parts:

- Knowledge Base, it stores expertise got from the experts in special area. Expert system is an intelligent system of this knowledge is the reasoning computer program using such knowledge.
- Reasoning machine, it has the reasoning capability which could derive conclusions according to the knowledge, rather than simply searching ready-made answers. It is a process module which implements a variety of tasks, carrying out various functions such as reasoning and searching.
- Database, it uses a framework of blackboard, and stores the data of user input every time. It also stores the middle data and final conclusion which is reasoned from

the inputted data of users. This database just describes initially the overall fault characteristics of the system, but with the opening of reasoning of knowledge base and increasing of the information feed-back information by users, it could give more comprehensive description of fault characteristics.

- Explaining system, it is used to answer the question from the users of the system, to give a explain about the process of getting conclusion or current state of the system, in order to enhance the credibility, trans-parency and learning capability of the system.
- Knowledge acquisition system, it imple-ments automatic knowledge acquisition of the machine, is implements self-learning of the expert system. It has the functions of adding, deleting, creating, updating, searching to the knowledge.
- User Interface, it translates the input in-formation of the human experts and users into the internal form which is acceptable of machines, and at the same time, it trans-lates the output information for the human experts and users into human understand-able form.

The shallow knowledge based reasoning has advantages such as intuitive knowledge expres-sion, unified form, strong modeling and fast infer-ence speed and so on. However, this method has great limitations, such as incomplete knowledge set, easy getting in trouble with unaccounted situation(Li M.& Wei Z.S.,2007).

The main problem of the expert system to face is the knowledge acquisition "bottleneck". On the one hand, for the specific fault, experts in the field can only know how to solve the problem but difficult to accurately describe their own knowledge and hardly say the reasons for this approach (Jiang Y., Chen X. & Yang X.,2002). On the other hand, different experts have the different knowledge and even conflict. The link between the symptoms observed of complex system and the corresponding diagnosis is really complicated. And the experience of human experts is hard to sum up into rules, especially a single rule. So this method is unfit for the system which has no experience of human experts for fault diagnosis. The application of machine learning method in expert system is expected to solve the problem of knowledge acquisition effectively.

Model-based knowledge reasoning and di-agnosing has more advantages on knowledge expression and knowledge organization than experience-based knowledge reasoning. It has more facilitation to get knowledge, and maintain it, and the consistency and completeness of knowl-edge base is easy to ensure, but the disadvantage is that the searching space is big, and the reasoning speed is slow.

The model-based (deep knowledge) diagnosis system requires definite relation between input and output in every part of the target, in the process of diagnosis, it generated the reason set which cause the inconsistencies between the output expected and the output actually created by the target(Li M.& Wei Z.S.,2007). Then in accordance with the first law in the field of target (with a clear scientific basis for knowledge) and the specific constraints in the target, it identify possible fault source using a certain algorithm.

Model-based diagnostic method adopts the model of operation and fault of the system that is to be diagnosed. The smallest diagnosis of system is computed according to the differences between the predicted pattern from the modules and the actual observed pattern. The principle is: firstly getting observed pattern via measuring tools, secondly obtaining the predicted pattern through the system input and the normal status of all components, then calculating the small-est confliction set according to the difference between the two patterns, finally computing the smallest diagnosis through the smallest conflic-tion set. The smallest diagnosis is an assumption

about fault components. Model-based diagnosis methods are classified fault model, causal model, stricture and behavior based model, diagnosing and reasoning model.

Model-based diagnostic method is independent on the diagnostic training examples and diagnostic experience of the diagnostic targets. It can diagnose multiple faults and explain the conclusion of fault diagnosis. But it is limited by the complexity of modeling and calculating for the large and complex systems. Furthermore, once the system is implemented, the systematic function is determined, and the system performance is unable to be improved any more through the use of successful or fail experience in the past(Jiang Y., Chen X. & Yang X.,2002). In recent years, the developing method which combined the experience-based method with the model-based method can solve the existing problems above-mentioned. Such as the integrated diagnosis model proposed by Galanti and Fink, causal level diagnosis model presented by Peng.

Diagnosis System Based on Neural Network

Neural network based intelligent diagnosis system focus on two researching aspects. One is using neural network as a classifier in the fault diagnosis from the point of application of pattern recognition, the other one is employing neural network as a forecaster for fault diagnosis from the point of dynamic forecasting (Wu J.P. & Xiao J.G., 1997).

Pattern Recognition Fault Diagnosis Neural Network

The fault diagnosis questions are regarded as pattern classification and pattern recognition in system. The superior performance of pattern classification is used to diagnosis. The model of neural network for fault pattern recognition can be classified into two categories according to their learning ways, namely supervised learning model and unsupervised learning model. The former mainly includes BP (back propagation) network and RBF (radial basis function) network, the latter mainly includes ART (adaptive resonance theory) network and SOFM (self-organizing feature map) network(Jiang Y., Chen X. & Yang X., 2002).

Supervised learning model has a very good ability for promotion in the fault pattern recognition. Trained BP network and RBF network have fast computing speed, and can be well used in the real-time monitoring and diagnosis (Yuan Z. R.,1999). However, these models have certain requirements on the distribution of the samples which is difficult to get in the practical application.

Fault Prediction Neural Network

The Fault Prediction Neural Network achieves the function of predicting fault in two main ways, one is to use the neural network such as BP network as a function approximation to fit some parameters of the target status. The other one is to predict fault by the establishment of dynamic model for the process or status parameters with dynamic neural network carrying feedback connection.

Using dynamic neural network to predict is a process of modeling the dynamic time sequence and this method fits the diagnosis to the dynamic system. The dynamic neural network has the ability of remembrance and association and classified into whole regression network and local regression networks with local feedback structure. All these networks have a common feature that their outputs not only depend on the current input, and the network itself has a corresponding dynamic structure. They all perform a dynamic prediction in the practical of modeling and forecasting of nonlinear dynamic system. But the dynamic neural network has a more complex structure than the feed-forward network, so it is more difficult in the sample training. Therefore, reducing the complexity of the network architecture reasonable and simplifying the learning algorithm of the

network are the important tasks in the practical application(Qian Z.Q.,2004).

For the purpose of improving the performance on the learning and intelligent diagnosis of the neural network, improving the neural network itself is main research issue. There are some improvements against the limitations of BP algorithm. There is also a new high-end neural network, named ellipsoid unit. The network forms decision regions of all pattern classes through the Gaussian distribution approximation, and reflects the distribution in the decision space clearer and more direct, so its classification precision is higher, and the training speed is rapid. The system has a very good capability of rejecting unknown pattern. So it has the clear superiority that it has making space of potential application(Zhou Z.D., 2004).

The idea of modular neural network is using a number of relatively simple network modules to run with a certain synergy to deal with a large-scaled problem. This strategy has received increasing attentions because of its effectiveness.

The Distinction of the Two Types

The most expert system of fault diagnosis systems is rule-based expert system which compiles a field of knowledge into a series of rules for storing and processing. The expert system is a symbols reasoning system which can solve complex practical problems by using of the knowledge and reasoning logic of human experts. It has lots of authoritative knowledge with learning ability, so that it can solve the problems that hardly can be solved in normal conditions on the use of certain strategy and reasoning upon expert knowledge.

Although this type of system can solve many fault diagnosis, it must utilize productive rules for the complex system diagnosis, so the expert system runs too slow to adapt in the requirements of real-time environment. In addition, the traditional expert system have the problems of hard acquiring knowledge, hard maintaining knowledge, narrow knowledge application and low ability of reasoning and solving problems. When meeting the fault that not existing in exist knowledge base, it cannot diagnose correctly. This system faces the problems such as "matching confliction", "mixture explosion", "unhelpful recursive" and other issues frequently(Qian Z.Q.,2004).

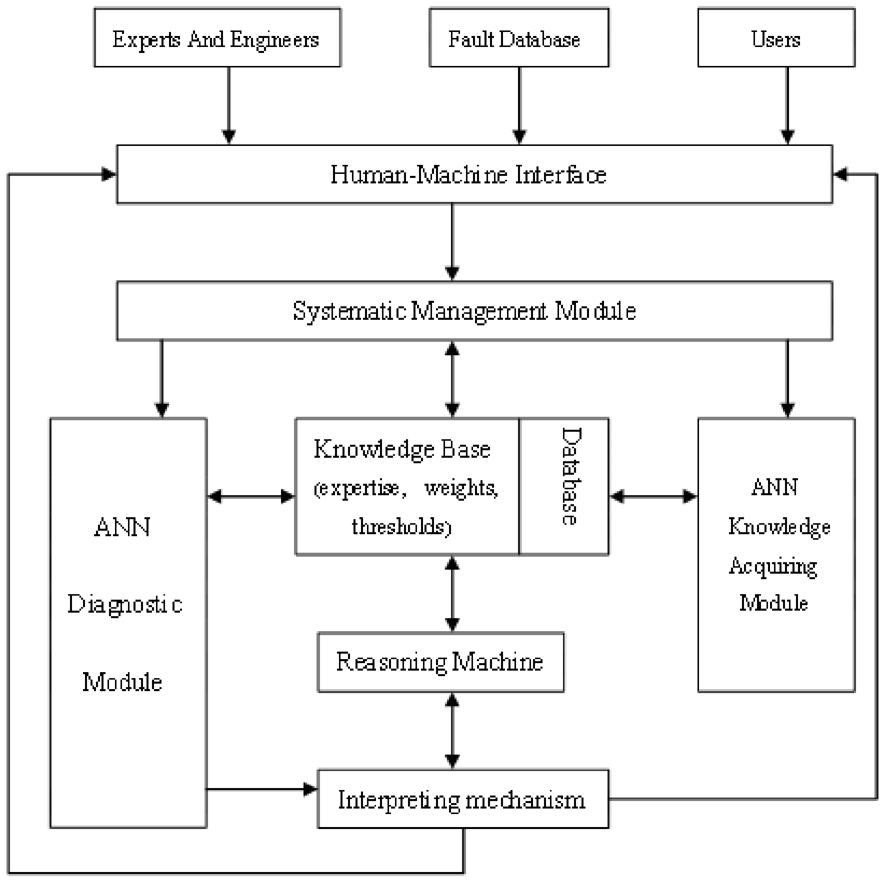
The neural networks are different from traditional expert systems which use symbols or data structures to represent knowledge (Wu J.P. & Xiao J.G., 1997). The knowledge implicit in the entire network rather than in a clear logic expression, that it is hard to know how it works (Raghunathan, Rengaswamy.,2000) . Once the result of the diagnosis system mismatch the predicted result, it is hardly to decided where the fault is. This is the constraint of neural network for knowledge management and fault diagnosis.

Combination neural networked with expert system, their advantages and shortage is complementary. Their combination has two ways. One is using neural networks to construct expert system, which is to transform the symbol-based reasoning of traditional expert system into the form of numerical calculating-based reasoning. It improves the efficiency of expert system and solves the learning problems. The other way is establishing a mix fault diagnosis system where the expert systems and neural networks co-exist, in this system they express and process different knowledge separately.

Mixed Methods of Intelligent Diagnosis

A single expert system has a weak ability on reasoning for the coupling relationship between causes and symptoms of fault and the relationship between the fault sources, but neural network is able to solve these problems with its parallel processing advantages(Li M.& Wei Z.S.,2007). A single neural network could not express the levels of diagnosis clearly, but expert system is handy for this. Modern fault diagnosis systems usually require a huge database for storing, querying,

Figure 2. Structure of mixed methods of intelligent diagnosis system



sharing and recovering from fault (Li M.& Wei Z.S.,2007).

Fault diagnosis is mainly performed by the neural network, whereas the expert system's main task is giving the cause of fault and countermeasures according to the results from neural network and the symbols of fault. Expert system is used to search for causes of fault, that is, implement the explaining function of the diagnosis system, and output the decision. The knowledge database is organized by database technology (Noel, Steven O' Berry.,2002).

Mixed method of intelligent diagnosis system structure is shown in Figure 2.

The models of system are introduced as follows.

1. HMI (human-machine interface) implements the function of interaction with the users, and transmits the real-time data of fault diagnosis to the next level by calling the system management model.
2. System management model has four functions.
   a. Creating the neural network by acquiring corresponding parameters table from knowledge database, and to transmit the data to the neural network.
   b. Translating the fault data into fault symptoms which can be identified by the expert system in order to complete reasoning.
   c. Determining whether to operate on the knowledge database or not.

   d. Storing the fault data regularly into the knowledge database by system management model when a new fault occurs, and to start ANN module to acquire new knowledge.
3. ANN diagnostic module completes the diagnostic process, get out the diagnostic results and transmits the results to explaining mechanism machine.
4. Knowledge database stores the rules of the expert system and the weight values, thresholds and structure knowledge of the neural network.
5. ANN knowledge acquiring module draws out sample data of global database, and starts a training of neural network, and then stores back the weights, thresholds, and structure knowledge to knowledge database.
6. The task of reasoning machine is to search for the causes of fault and countermeasures in the knowledge base in according to symptoms of fault and results of neural network, and the other function is to achieve the user's fault information enquiries.
7. Explaining mechanism is implements by expert system, which calls the reasoning machine to give out the reasons for fault and fault countermeasures in according to the symptoms of fault and the results of neural network.
8. Fault database stores the detection data from the detection device and updates them regularly. Different fault data from different modules will be stored separately in order to carry out fault diagnosis smoothly.

Knowledge acquisition is not only from the background knowledge of objects and expertise, but also from the training of neural network.

Knowledge expression is performed by productive rules in the expert system which expresses background knowledge of the objects and expertise, and this knowledge is organized by database. Weights and thresholds of neural network are stored implicitly in the neural network structure. This knowledge is also organized by database in the form of parameters table.

The system select the target sub-modules in accordance with user's selecting, and then selects fault data of the modules from fault database to start the diagnostic process while the system management module translate the fault data into symptoms of fault with a certain way and submits them to the exert system. After the diagnosis of ANN diagnostic module, the neural network outputs the result of diagnosis process to the explaining mechanism module. When a new situation comes out, the neural network will not diagnose it clearly, and the fault will get other procession through the judgment of output of neural network.

The neural network has corresponding output range with the types and extents of fault. When the diagnostic result is out of this range, a new situation will be recognized by the system. At this point, the system management module will store the sample data of new fault into knowledge database, and will prompt the user whether to start knowledge acquisition when the fault data accumulated to a certain level(Zhou Z.D., 2004). The users or the system designers draw out the fault data, choose samples and select a suitable structure of neural network with the message from system management module in order to train the neural network correctly. After training of the network, the knowledge acquired by the ANN knowledge acquisition module will be stored into knowledge database, and this is a complete process of knowledge acquisition.

According to fault symptoms and diagnostic results from the ANN module, expert system module will launch the reasoning machine, and search for the causes of the malfunction and fault countermeasures, then put these information onto the human-machine interface in order to show the useful information to human users.

If there are no causes and countermeasures of this fault in the knowledge database, the explaining agency will prompt the user to do some

further deal, such as increasing the causes and the countermeasures.

Intelligent diagnostic expert system based on neural network has more advantages than traditional expert system, however, there are also some weakness with the current neural network expert system(Jiang Y., Chen X. & Yang X.,2002). Mostly one is that, the neural network just simulates the intelligence and activities on the level of feelings of human, and hence people are trying to research out a integrated intelligent system that combines both symbol-based reasoning and numerical calculating-based reasoning to simulate human thought process more better.

Diagnosis System Based on Case Reasoning

Case based reasoning (CBR) can solve new problems through amendments to the successful outcome of similar problems. It is a method that uses the experience and cases of past as a guide to solve new problems.

The principle of case based diagnostic method is that according to the characteristics and symptoms of the target, it will find out the most similar case to match the problems which the target have, and then the result of the case had found out will be modified as the diagnostic result of the target.

This method is able to be applied to the areas where theorem can hardly be expressed in a form of rules but in the form of cases. The critical thing to implement this method is to establish an effective case index mechanism and case organizing mechanism.

The CBR diagnostic method mainly contains the following aspects(Jiang Y., Chen X. & Yang X.,2002):

1. **Case Expression.** It can be expressed with symbols of fault, the causes of fault, and the occurring environment and a series of characteristics.
2. **Case Index.** It can find out relative cases from case base quickly and accurately. Case indexing technique has primary three kinds, namely, the last adjacent law, the induction law and the knowledge-based law.
3. **Case storage and Case retrieval.** It must store the new cases and organize the characteristics of cases.
4. **Case Modification.** If there is better expression for a case, it will modify the case.
5. **Studying and Summarizing.** It can get the collection of the knowledge of diagnostic process and use the knowledge accurately.

The CBR method is good at expressing difficult causal relationship, but it also has its limitations (Sallans B., 2002). The traditional case-based diagnostic method is difficult to interpret the link between cases, and it is very time-consuming the retrieving in the large-scale case base, in addition, it is difficult to decide which symptoms and their weights to choose.

When the searching goes on, there may be some optimal solutions lost, or if abnormal symptoms come out, it may be misdiagnosed or missed because that it is unable to find the best match.

The CBR method is difficult to deal with the consistency test (restriction of characteristic variables) for the modifying of cases, and also difficult to explain the results of diagnosis.

An effective solution to these problems is to integrate other techniques in the CBR method, for example, the fault tree analysis, rule-based reasoning and artificial neural network methods.

Fuzzy Logic Method

The fuzzy theory based diagnostic method come out because of the uncertainty of some systematic status, and incomplete information acquisition.

The diagnostic methods based on the fuzzy relationship and synthesis algorithm usually determine fault and uncertain relationship between the symptoms using the appropriate membership

function and fuzzy relationship matrix (Wu J.P. & Xiao J.G., 1997).

This method is first to make the causal relationship matrix R between sign of fault and types of fault, and second to establish the fuzzy relationship equation with the fault and the signs. The expression is F = S °R, where the F is for the fuzzy fault vector, the S is for the fuzzy signal vector, and the operator " °" is for the synthesis operator.

During the process of diagnosis, the first step is to test the parameters of status of the target to be detected, that is an extraction of characteristic parameters vector matrix X.

By solving the relationship matrix equation Y = X °R, the fault vector Y of the status to be detected can be seized, and the result of diagnosis is coming out under some certain principles of judgment of diagnostic process.

The critical step for implementing the fuzzy diagnostic method is to choose the appropriate membership functions and representative characteristics. The determination for the membership function has mainly three methods such as fuzzy statistical testing method, the weighted statistical method, the binary comparing sorting method and dynamic signal analysis method.

The most important feature of fuzzy logic system is that, its fuzzy rule database can be structured using expert knowledge directly. So it can make full use of the knowledge and experience of human experts, and deal with the experience effectively. In addition, a good designed fuzzy reasoning system can be arbitrary in a non-precision approach to a linear function with good performance(Jiang Y., Chen X. & Yang X.,2002).

This approach can resolve the uncertain questions in the diagnosis, but the knowledge of fuzzy diagnosis acquires difficultly, specially, the fuzzy relationship between fault and symptoms can hardly be determined. The system based on fuzzy logic method has a weak learning ability, so that it may misdiagnose or miss fault. In addition, because of ambiguous language variables is identified by membership values, the problem of implementing, is how to achieve the conversions between linguistic variables and membership values (Wong, S.V. & Hamouda, A. M. S.,2002).

The emergency of fuzzy theory, fuzzy logic system and fuzzy neural network provides an important theoretical methods and tools for solving the diagnostic problem of complex systems. Specially the FNN (fuzzy neural network) technique has been resulted in because of the integration of fuzzy logic system based on the fuzzy theory and the neural network(Qian Z.Q.,2004).

The FNN technology is the integration of the fuzzy logic and the neural network whose essence is the double simulation of human brain structure and thinking function. That is, to simulate the hardware topology of brain network and the software feature of fuzzy information processing function at the same time. It can deal with expert knowledge and experience and enhance the judgment ability of the system through self-learning. It can optimize the fuzzy rules and membership functions or fuzzy decision-algorithm in the fuzzy logic system.

REMOTE INTELLIGENT DIAGNOSIS AND MAINTENANCE SYSTEM

The remote long range integrated system is the organic integration of remote monitoring subsystem, diagnostic sub-system and maintenance sub-system using network technology. It can change the distribution and structure of the entire system flexibly according to the actual situation of the equipment operating in the process(Zhou Z.D., 2004). It can also solve the problems that may arise of the equipment in the shortest possible time, so that minimizing the cost of equipment and maintaining the operation of equipment stably and effectively.

At present, there are several structures being widespread concerned and in-depth studied in the remote intelligent fault diagnosis area.

The C/S (Client/Server) Architecture.

The client side includes remote workstations and local workstations, its main functions are processing in accordance with certain application logic and communicating with other clients and interactive database system(Li M.& Wei Z.S.,2007). The main functions of the server are operating on the database in accordance with the request of the clients and giving back the results to the clients.

In the C/S structure, the server provides services to clients, and they communicate with each other with message mechanism to complete services request and transmission of result information of diagnosis.

There are two main characteristics of the C/S structure.

1. **User Transparency.** The users have not to know the specific location of the server, but they can make full use of the sources which the server have managed and all kinds of services available provided by the server.
2. **Closed Service.** The service mechanism within the server is opaque from the foreign region. It interacts with the outside world only by the messages.

Traditional network-based diagnosis system almost uses this C/S architecture.

The B/S (Browser/Server) Architecture.

The B/S structure is an extension of C/S structure. Because of its multi-layer architecture, it has more advantages than the former. First of all, it is an open cross-platform system allowing the users to access a number of application servers with the browser. Secondly, the system uses the standard browser as the graphical user interface (GUI) in order to simplify the operation of the users.

This structure generally includes the browser on the client side, the remote diagnostic servers and network diagnostic resources(Li M.& Wei Z.S.,2007). Diagnosis process is as follows: First, the users choose HTML pages in accordance with its provision of services when the users visit the remote diagnosis center site server through the browser. Second, users submit the collected data documents to the web server in accordance with the types of the fault sign, and start the reasoning machine through the middleware interface, which can get out the result of fault diagnosis and the maintenance advices with corresponding knowledge base and diagnostic rules. Finally, the web server organizes the diagnostic result and maintenance advice into the form of HTML for the downloading request of the users.

The features of the B/S structure are introduced as follows.

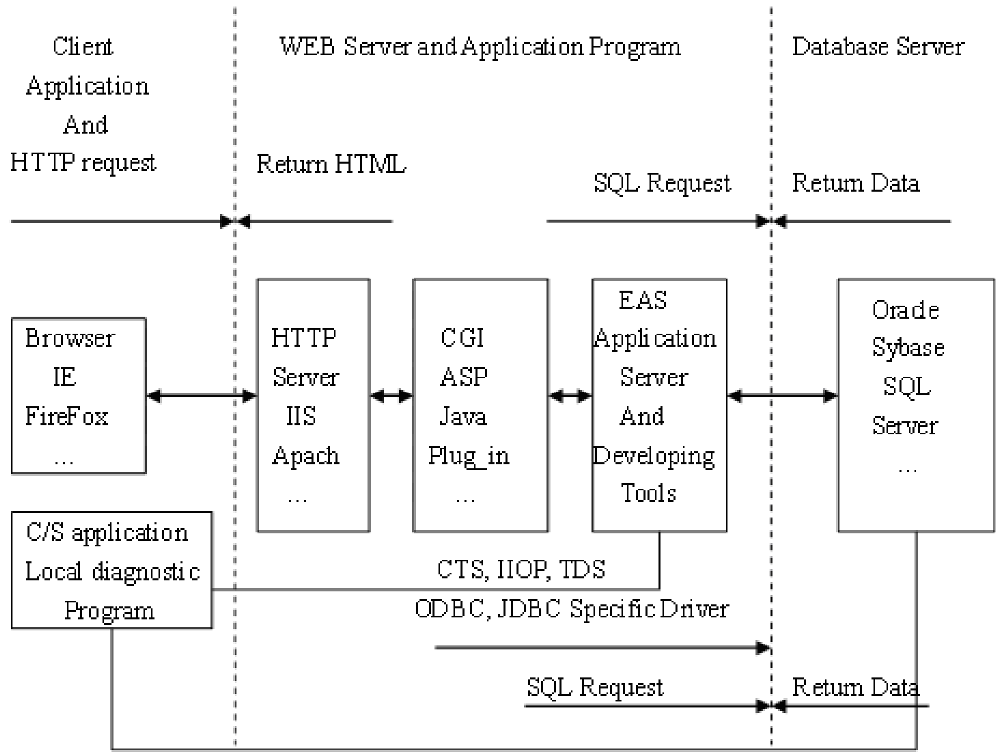
1. Sharing the network diagnostic resources. It will be a rich diagnostic database.
2. Could achieve the diagnostic collaboration of human beings and machines and other elements of diagnosis.
3. Open diagnostic system using web database, which updates and improves the knowledge base and diagnostic rules constantly.

The Mixed Multi-Layer Mode Architecture

The whole system is divided into three layers: the user layer, the middle application layer and the database layer

The middle layer is the core of the whole system, which has the functions of dealing with the business in accordance with application logic, contacting with the user layer and the database layer. The database layer is responsible for the data request from the middle layer, and its main function is defining, maintaining, visiting and updating the data in it. The user layer provides the interface between users and the system, and it deals with the interaction of the system and the users.

Figure 3. Architecture remote system framework based on mixed multi-layer



The three layer structure of this system have enough openness and flexibility so that it make the client simple, and transfer the work of development of management to the servers(Li M.& Wei Z.S.,2007). In this structure, the application logic is relatively concentrated. From the point of developer, it can easily adapt to the application needs of the first two architectures above, and the developer can use all the techniques such as CORBA, Java Bean, COM/DCOM and so on.

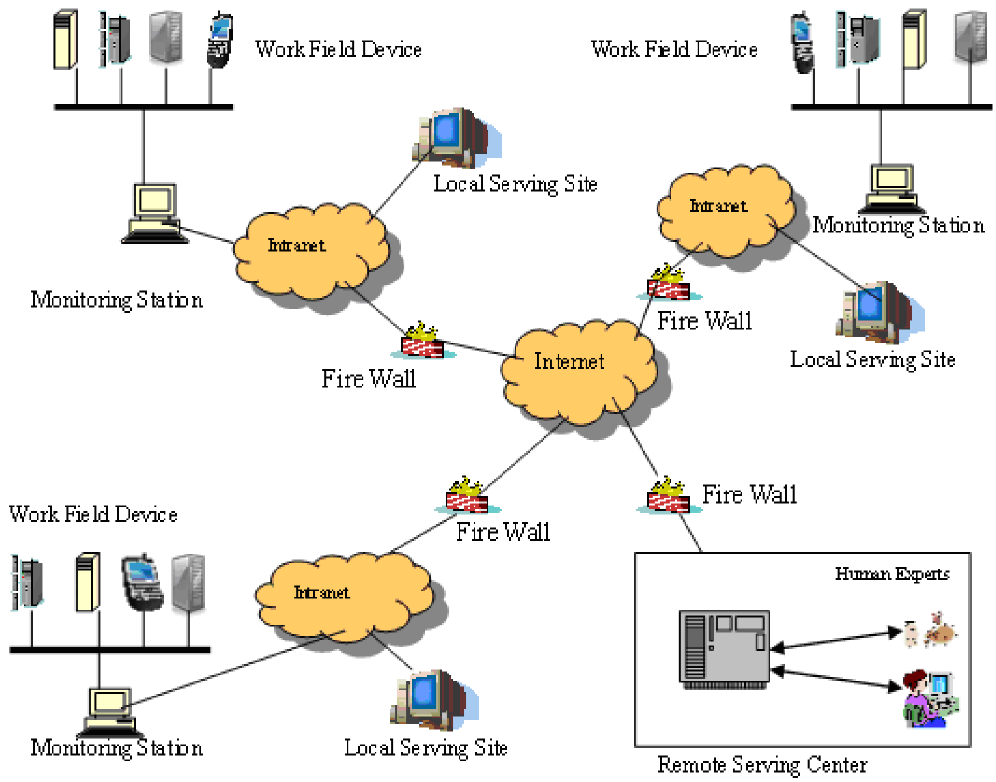
The entire system architecture model of the internet-based remote diagnosis system is shown in Figure 4.

The system is divided into three main parts: the working spot monitoring systems, the local serving sites and long-distance remote serving center sites(Jiang Y., Chen X. & Yang X.,2002).

The results of spot monitoring system is the basis of local serving site and remote serving center sites, which is responsible for collecting the various data in the course of the operation of equipment. This data and the situation is the main basis for the fault diagnosis and the maintenance of equipment. At the same time, this sub-system also has the capabilities of handling with the emergency situations and the interactive ability with local sites.

The local serving sites judge on the operating situations of the equipment with the results of spot monitoring system to decide whether to take down and maintain the equipment. If it has diagnosed an occurred fault, it will give out the results and get the support of remote intelligent fault diagnosis system when it is necessary. It is responsible for arrangements for guiding the work of maintenance workers. The relevant records must be stored into information service resource database, and they are the references for the future work.

Figure 4. Architecture of internet-based remote integrated system



The long-distance serving center sites have better technical strength and more information service resources as well as a huge expert system database, which provide wide-area sharing of resources within the platform, and establish the collaborating mechanism of multi-diagnosis system and sharing mechanism of resources. They also provide on-line support for the local on-line diagnostic system. The core resources are the diagnostic experts and the diagnostic knowledge base. The knowledge base on the one hand provides corresponding knowledge for the equipment's diagnosis. on the other hand, it gets knowledge from local on-line diagnosis sites which is refined, extracted to enrich the knowledge base.

From a point of view of system functions, the remote distributed intelligent fault diagnosis system has three components: local on-line fault diagnosis system, remote intelligent fault diagnosis system and diagnostic experts.

The diagnosis experts have played an important role of the fault diagnosis system in the systematic operating. They are mainly collaborating with remote diagnostic system for providing service for the users. In addition, the diagnostic experts add the diagnostic knowledge from diagnostic systems distributed at all places to the knowledge database of remote diagnostic system which has been analyzed, sorted and refined.

The remote intelligent fault diagnosis systems collaborate with the diagnostic experts and they interact with the local on-line diagnostic system, collect the fault knowledge from all places and support the local on-line fault diagnostic system. They also sum up the features of the fault from local equipment, and extract the essential char-

acteristics of the fault knowledge in accordance with operating mechanism of equipment.

The local on-line fault diagnostic systems face the devices more directly supporting them with on-line diagnostic service. Their services are provided to the respective targets concentrated in the area which it is belonged to. In addition, through the interaction and collaboration with the remote intelligent diagnostic system, they can acquire corresponding support and submit local diagnostic knowledge and experience to the remote center sites.

With the above introduction and analysis for the systematic structure and functional model, the workflow of the system is induced.

After the local serving site getting the conclusion according to the analysis for the monitoring data and the working field situation, it determines to take one or more measures in the following three aspects:

1. If there is no fault occurred, then it continues to monitor the situation and guide the maintenance.
2. If there are some faults occurred, then it starts the diagnostic modules, and develops maintenance plans, arranges maintaining work in accordance with the results of diagnosis and the corresponding information resources.
3. If there are technical difficulties to resolve the faults occurred, then it will seek for help with the long-distance remote serving center sites.

When the remote serving center sites receive the requests from local serving sites, they deal with the requests using corresponding expertise and information resources in accordance with remote monitoring techniques and parameters submitted. If there are also some technical difficulties, it will submit diagnostic request to the remote human experts, and make arrangements for video conference to organize tripartite force for solving the problem jointly.

MULTI-AGENT-BASED INTELLIGENT DIAGNOSIS SYSTEM

Agent-Based Diagnosis

From the learning about the theory of the chapter 3, we know that, the Agents have perceptual ability, resolving ability, communication ability with others. So construction with Agent to imitate human expert's logic in the process of fault diagnosis is possible. In the MAS diagnosis system, the diagnosis mission was decentralized to a number of agents for processing, by using the cooperation and competition within the agents to complete the diagnostic mission.

Fault diagnosis includes diagnostic reasoning part and knowledge studying part. Relative independent agents could be constructed according to each part, but they are a unity logically. According to the common models of reactive agent and cognitive agent, the diagnostic agent here is mainly composed of fault diagnostic machine, knowledge learning machine and diagnostic knowledge database.

Attendance perceiving machine perceives the information of attendance from environment, and receives the status information of the relative target. The fault diagnostic engine receives the information and diagnoses for the target using the diagnostic knowledge in the knowledge base. Then the results of diagnosis are outputted by the outputting machine or are transmitted to the learning machine through the communication mechanism.

The knowledge perceiving machine perceives new information from the environment and transmits it to the knowledge learning engine which adds the new knowledge into the diagnostic knowledge base, or does some adjustments on the old knowledge that is not appropriate to enhance the capability of the fault diagnostic agent. It also submits the knowledge and corresponding information to the outputting machine in order to affect the learning environment.

The diagnostic knowledge base is the knowledge storing center of the diagnostic agent. The abstract thinking agents store the knowledge in the form of symbols, but the specific thinking agents store the knowledge in the form of neural computation.

Model of Multi-Agent Diagnostic System

Macroscopic the multi-agent system is a simulation of the society, but microscopic it is the development of different entities with different functions for the use of different applications.

From the learning about the theory of the chapter 3, we know that, the Agents have autonomy, responsiveness, initiative, sociability. So we can use a number of agents to make a MAS diagnostic system, in order to get the agents cooperating for parallel processing.

From the learning about the theory of the chapter 5, we know that, artificial neural network has simulated the human brain. The neural network has powerful learning and parallel processing ability. The distributed information storage mean of the neural network, has provide a new way for the knowledge storage, acquisition and reasoning.

The MAS (multi agent system) diagnostic system is constructed by BDI agents. The diagnostic system has the following main sections(Zhou Z.D., 2004).

1. **Monitoring Agent.** It monitors the machine and collects the state information of the machine. When there is a abnormal state occurring, It sends a message for attendance to the management subsystem and provides the initial data to the subsystem, and then it notifies users about the anomaly.
2. **Systematic Management Agent.** After the receipt of the attendance request, it requests the appropriate diagnostic agents for fault diagnosis, and it provides an interface to HMI for submitting the final conclusions and corresponding explanations. It completes tasks such as the distribution of tasks, agent communication and coordination and so on with some certain coordinating strategy rules.
3. **Blackboard Agent.** It gets all the diagnosis results of different diagnostic agents, it is also responsible for the tasks of interactions, recording the needed information of the agents, making decision conclusions and making the information sharing within all agents.
4. **Conflict Resolution Agent.** When the blackboard agent has detected the conflict, a notice will be sent to this agent. It monitors the blackboard agent, and starts the task of conflict resolution when there is information conflict occurring between users.
5. **Diagnostic Agent.** This agent reads in the necessary data from the monitoring agent. After diagnosis process completes, the results of diagnosis will output to the blackboard agent. Then it will request the HMI agent to display the conclusion at the screen.
6. **Human-Machine Interface Agent.** It is the interface of the whole system. It is responsible for providing the interface to show the sampling data of monitoring subsystem and the conclusion from the management agent. When the system needs users for participating in the diagnosis, the users can operate through the interface.

Implementation of the BDI MAS System

In the diagnosis process, the monitoring agent will send the monitoring data to the HMI agent for displaying and handling. When it has detected anomaly, the attendance request will be submitted to the management agent. Then the management agent will identify and classify the observed phenomenon and the parameters and character-

istics got from the monitoring subsystem. The classified information will be distributed to the corresponding agent and submitted to the symptom database for re-identifying and re-classifying in order to form the initial data for starting a certain agent(Zhou Z.D., 2004).

The diagnostic agent refers to the knowledge base using the reasoning machine, and gets out the ultimate conclusions and outputs the information to the blackboard agent for sharing. During this period, the agent also can request to confirm the fault or the exclusion of the fault.

Each agent does not have a complete ability of solving the problem comprehensively and all-round knowledge. If the problem could not be solved by a certain diagnostic agent, it will be sent to the blackboard agent and be re-distributed to appropriate agent for solution. The middle data from the last agent will be formed into the initial data for the next agent. This is the so called joint agent diagnosis(Li M.& Wei Z.S.,2007).

At the same time, the human experts could participate in the diagnosis and decision-making process through with the HMI agent, forming a joint human-machine fault diagnosis(Jiang Y., Chen X. & Yang X.,2002).

This model takes full advantage of the awareness of the BDI agent with the environment, improves the diagnostic ability of the system, while reducing the complexity of the structure of the whole system, and it also enhances the effectiveness of the state monitoring.

Mission Decomposition of the Fault Diagnosis

The task decomposition is the crucial step to affect the effectiveness of the system.

There are four main methods of distributed task decomposition, the method according to the spatial distribution of the applying system, the method according to the functions of the sub-systems, the method according to the inputting styles and the method according to the outputting styles.

The principle of the task decomposition requests clear purposes of sub-tasks, least coupling between sub-tasks to simplify the communication of the agents completing the diagnostic sub-tasks. On the coarse- grained level, the structure based decomposition strategy is usually used, but on the fine-grained level, fault based decomposition strategy is usually used. Through the distributed decomposition of the diagnostic problem, a clear hierarchical task tree will be created.

Strategy of Diagnosis Controlling

The diagnostic sub-tasks at the same level can be performed parallel, but the super task must be performed after the completion of its all the sub-tasks.

1. **Parallel Controlling Strategy.** Each task has a thread which is started by the super thread. Thus, the diagnostic system starts the global diagnostic task thread from the top of task tree, and then starts the sub-task threads in the lower lever until the bottom. This strategy has the advantage of fast implementation, but the disadvantage of excessive consumption of the systematic resources.
2. **Serial Controlling Strategy.** This strategy is that, the diagnostic system starts from the global task on the top of the task tree, implements every diagnostic sub-task in a single thread with the DFS principle. This strategy has the advantage that the implementing route is simple and the systematic resource consumption is little, but the disadvantage of slow implementing speeds because of serial operation.
3. **Hybrid Controlling Strategy.** This strategy is that, on the coarse-grained level it uses the parallel method, but on the fine-grained level it uses the serial method. This strategy can make the diagnostic system to achieve a reasonable balance between

implementing speed and systematic resource consumption.

CBR with the Coordination of Multi-Agent

The CBR is based on the following two principles(Li M.& Wei Z.S.,2007):

1. The similar problems have similar solutions.
2. The similar problems will happen again.

The coordination of the agents could be achieved by the CBR technology. Inside the agent, CBR is used in the process of reasoning, and the results will be the basis for coordination between agents.

The system used numeral symbols to express the information of the diagnostic process. For example, the fault sources are figured by the numeral ID of the sensors, the diagnostic algorithms are also figured by the numeral ID. That is, all the keys in the cases are number.

For example, if there is an anomaly occurred, the data of sensors 1,3,4 is abnormal, and in this diagnosis, the algorithm 2 (such as neural network algorithm) is the best, then this case will be expressed as 13402 (where the 0 is a separator). It will be stored as a diagnosis case.

In the multi-agent based fault diagnostic system, the content of case databases is different with different agents. For the mission assigning agent, there is successful diagnosis stored for different functional agent, but no non-optimal diagnosis. But for the functional agents, there are respective unsuccessful examples stored for them. The purpose of this is to improve the diagnostic efficiency and to reduce the burden of the mission assigning agent(Li M.& Wei Z.S.,2007).

With the increase of the diagnosis, the number of the cases has also been enriched, and the link between faults and certain agents will be gradually strengthened.

CASE STUDY

The Agent technology is currently used in the area of computer science, fuzzy control, control of production widely. The Agent technology has been adopted to deal with complex issues in practice. The environment can be perceived through the sensor of the agent, and the agent can play a actor role to react to the environment. To design a fault diagnosis system for distributed fault diagnosis and respond to the abnormal situation is a problem worthy of study.

Application Background Analysis

We assume that there is such an application scenario, there are a number of digital devices in a workshop, and they need on line monitoring and maintenance. A traditional approach is to make a maintenance engineer to carry out daily inspections of the workshop, when a device had a fault, they do their maintenance, this way is a passive maintaining method. The monitoring system used sensors or camera equipment on the plant to monitor and control all equipment, the maintaining engineer managed all the equipments on the monitoring spot, this system had a ineffective performance. So, the agent technology can be introduced into for higher efficiency.

The system has a number of computing nodes, behaving the function modules of the system, each node has a running platform for agents, the platform can support collaborative diagnosis. Part of function agents carried out monitoring task for devices to collect corresponding data from the devices, part of the agents carried out diagnosis task for devices using the abnormal condition data collected to make appropriate diagnostic conclusions. Some agents provided support for function agents to share resources, and managed the information of all the agents.

In the process of diagnosis, the human experts had involved, so there are some agents provided

human-computer interface, and managed the communication with other systems on the Internet.

Based on the above analysis, the monitoring and diagnosis system should have the following functions:

1. Using a variety of graphics and real-time interface to reflect the operational status of equipment.
2. Using a variety of data storage means to record the parameters of running conditions.
3. Processing and diagnosing the abnormal parameters in time, and make the appropriate warning and take maintenance for equipment timely.
4. High security, stability and good communication capability.

Solutions

In the diagnosis system, each equipment which need diagnosis and maintenance to safeguard had a computing node which had a IP address. Above it there is a agent running platform. Data transferring through IP net, and the whole distributed system is integrated by using the Internet, so a monitoring machine can be placed in a room at other places, which had more computing resources and could carry more complex agent and implement HMI task.

The working statuses of digital devices had been collected by front-end computing nodes, a high-performance embedded computer for example. On the node, we could ran the agent platform, and start monitoring agent and parameter extracting agent for example. The fault diagnosing agent and managing agent or more complex agent had been placed on the remote node which had more resources, user could achieve remote starting and controlling for each sub-systems by the support of HMI agent and managing agent. Diagnosis task would been decomposed into sub-task which would been sent to diagnosing agent to carry out, and the diagnosis results would been presented to the users. Some of the agents would been in the state of hibernation in some conditions.

The Linux operating system had been applied on the embedded computer in a wide range. Linux had got a lot of improvements for transplant in order to adapt to the environment. JVM would been transplanted onto Linux easily, which is a abstract computer, owning its own instruction set and use a different storage area. The lightweight KVM which belong to J2ME could provide a complete Java Runtime environment. An embedded Agent node had been designed according a embedded system actually. It should have the following functional modules: power module, processing module, data acquisition module, data storage module, communication module and warning module. As shown in Figure 5.

System Modeling

The framework of the monitoring and diagnosis system had been shown in the Figure 6.

The normal implementation of the system is clear. In each specific processing of diagnosis, according to the cyclical request of the HMI agent, the monitoring agent send the collected data to HMI agent for displaying and processing. when the monitoring agent had detected abnormal status, it would notify the HMI agent to request diagnosis which would be controlled by managing agent. The managing agent then would start parameter extracting agent, and then the monitoring agent would send parameter data up to it. The parameter extracting agent would submit useful feature data to the blackboard agent, and then would start analysis agent to identify the data. This data would be assigned to appropriate diagnosis agent which would use it as initial diagnosis data. Diagnosis agent look up to their own knowledge base for reasoning to obtain diagnostic conclusions, and then it would output to the blackboard to share with other agents. Decision making agent would confirm the fault and the exclusion of the fault. In the meantime, human experts could request

Figure 5. The structure of embedded agent node

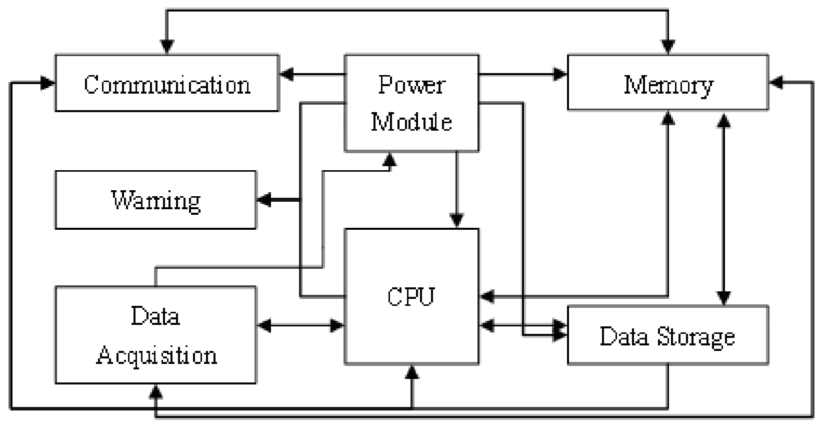


Figure 6. The framework of the monitoring and diagnosis system

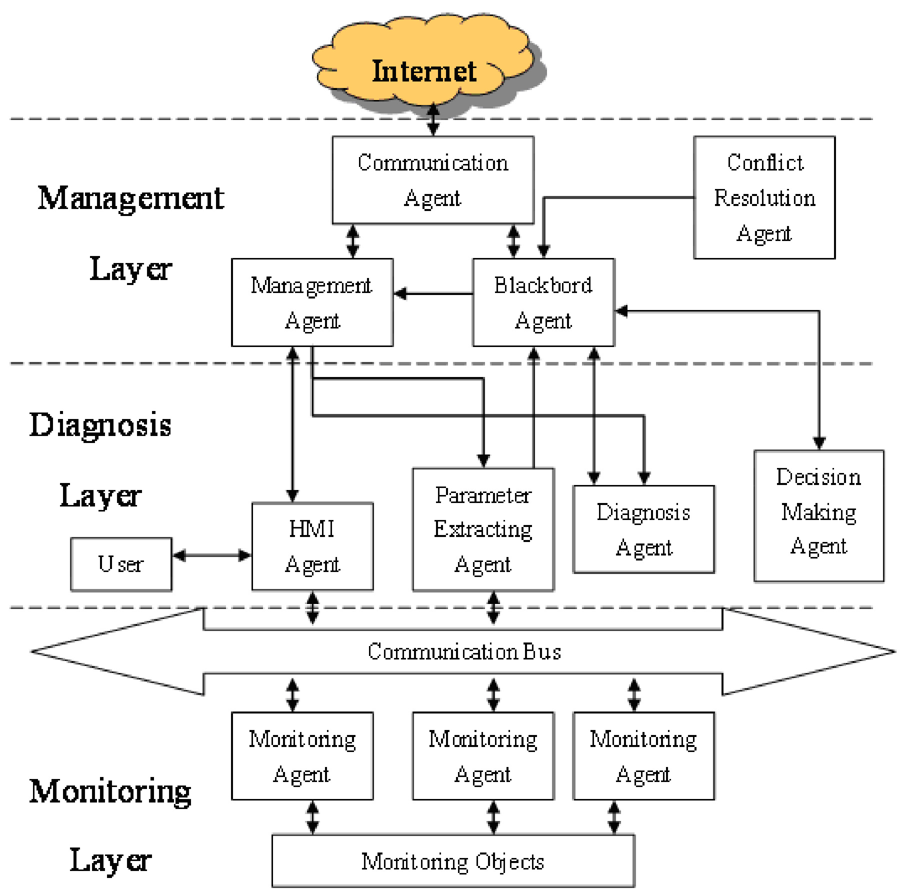
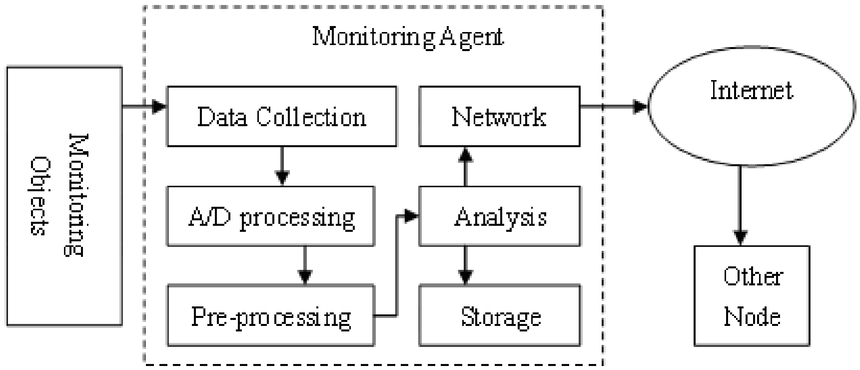


Figure 7. The basic modules of the monitoring agent



to managing agent for registration through HMI agent. Humen experts could participate in a joint diagnostic process and the decision making process to form a joint human-machine diagnosis.

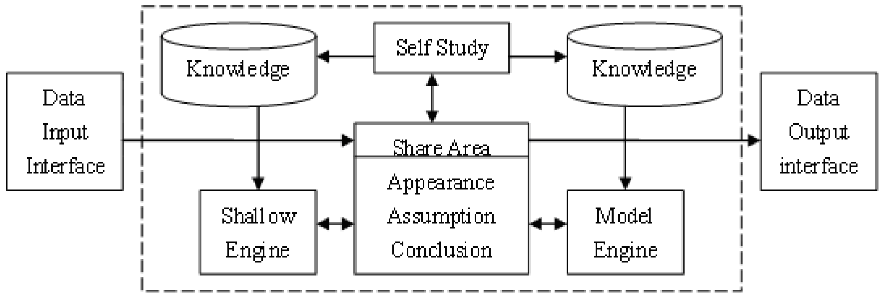
Specially, the realization of each agent contains the basic construction, message handling, such as sending and receiving. The basic construction had three steps: first, to inherit from the base agent. On the JADE platform, the class-jade.core. Agent is the base of all Agents, and implementing the method called setup would create an agent; second, to resolve startup parameters by using optional method named getArguments; third, to register the ontology and start agent's behavior in the necessary method setup. The simple code of the creation of a basic agent has been shown as follows:

```
import jade.core.Agent;   //im-
port related package
public class DiagnosisAgent ex-
tends Agent
{
      //print a welcome sentence
   protected void setup()
   {
      System.out.println("Hello!
Agent" + getAID().getName());
}
}
```

In the environment of JADE, the message had been corresponded to ACLMessage class. The communication had a asynchronous mode, each agent had a queue which is assigned by the platform. A simple code of message sending and receiving has been shown as follow:

```
1. A send message
// create a message
ACLMessage msg = new
ACLMessage(ACLMessage.INFORM);
msg.addReceiver(new AID("B",AID.
ISLOCALNAME));
msg.setLanguage("Chinese");
msg.setOntology("diagnosis_on-
tology");
msg.setContent("Device A running
abnormal");
// send message
send(msg);
2. B receive message
//receive a message
ACLMessage msg = receive();
if(msg != null)
{
   // deal with the message
```

Figure 8. Structure of diagnosis agent



}

The design of primary agents has been described.

Monitoring Agent

The monitoring agent has an on-demand transmission mode for status data. The embedded agent node would pre-process the data after data collection had completed and analyze the condition, then it would put data on to network for other nodes selectively. the Figure 7 has described the basic mode of the monitoring agent.

It could sense the information of the running status of equipments automatically, capture information abnormally for diagnosis agent.

HMI Agent

The HMI agent is the main channel of human-machine interaction. It is responsible for displaying the data from monitoring agent and blackboard agent. When the participation of human experts had been needed, it also provide a operating interface for human to implement interaction. The HMI agent print out the record of the real-time conditions of the monitoring objects, and it also would give out the alarming for responding abnormal status.

Parameter Extracting Agent

The parameter extraction should be important in the diagnosis process, it directly affects the calculation of the entire diagnostic process, as well as diagnostic performance of the system. According to the requirements of the parameters, it analyzes and arranges the original data which would be send to by monitoring agent. The parameter would be send to blackboard agent for restoring, and it could be the initial accordance of appropriate diagnosis agent.

The original high-dimensional parameter vector would be expressed by low-dimensional vector by the parameter extracting agent, and it would be organized for fault diagnosis.

Diagnosis Agent

The information for diagnosis had been mainly from parameter extracting agent, and the diagnosis agent would explain the condition of equipments abnormally and diagnose the fault. It had been started according the data restored in the blackboard which is transfer by the blackboard agent. After the fault diagnosis process, the result of diagnosis would be sent back to the blackboard. The inner structure of the diagnosis agent has been shown in the Figure 8.

At the beginning of the process of fault diagnosis, the diagnosis agent would start the shallow knowledge inference engine to produce some

hypothesis of the failure which had been concerned most likely, and the assumptions should be confirmed by the deep knowledge reasoning engine so as to arrive the conclusion of the failure. The shallow knowledge inference engine make diagnosis by using artificial intelligence methods, and the deep knowledge reasoning used structural knowledge and the record knowledge of fault which had been the historical basis for fault diagnosis.

FUTURE OF INTELLIGENT DIAGNOSIS

The knowledge based diagnostic system has a trend of from the rules-based system to the hybrid model based system, from human expertise inputting to machine learning, from static diagnosis to real-time diagnosis. In addition, there is more research on with the network based remote distributed diagnostic system and has a good prospect.

Integrated Diagnostic Technique

Every diagnostic method has its own advantages and limitation. To integrate a wide range of diagnosis techniques effectively for enhancing the comprehensive performance of the diagnostic system, is the absolute trend of the development of intelligence (Golobardes E., Lloro X., Salamo M.& et al.,2002).

The diagnosis for the actual target is complex, and it is hard to express the fault knowledge comprehensively with a single knowledge expressing method. Thus integrating a variety of methods can express the knowledge of the fault diagnosis for the target better.

The research on the integrated diagnostic technique should consider with the characteristics of the target, the advantages and the limitations of the diagnostic methods, and then take the appropriate integrating strategy. A special strategy is to integrate the method based on rules, the method based on neural network and the method based on case into one architecture, so that to make full use of the advantages of these diagnostic methods to improve the performance of the system.

Human-Machine Cooperation—from Single Work to Remote-and-Distributed Diagnosis

The existing equipment diagnostic systems mostly face to single machine, or single class of machine. They have a weak ability of scalability, flexibility and versatility. The systems are separate from each other, so that they can hardly exchange or sharing information, resulting in a tremendous waste of resources.

Nowadays, many large complete sets of equipment or institutions are composed by different sub-systems distributed in different places. Correspondingly, diagnostic information would be transmitted and exchanged between the sub-systems (Zhang Q.B.,1997).

Therefore, setting up a remote distributed intelligent diagnosis system which use a number of diagnostic machine to cooperating in a diagnostic process is a better resolution for complex diagnosis (Stanek M., Morari M.& Frhlich K.,2001). In addition, it allows remote human experts in different places to participate into the diagnosis, and it also allows lots of equipments to share one diagnostic system. By this way, the diagnostic cases will be accumulated rapidly to enhance the system effectively.

Behavior-Based Dynamic Intelligent Diagnosis

From the point of dynamic physics of the system itself, the research on the dynamic intelligent diagnosis will make the solution of the problem more effectively. It is good for taking full use of dynamic system information, describing the systematic status, diagnostic knowledge and reasoning process better.

Traditional knowledge acquisition of the expert system is difficult. But behavior-based intelligent diagnostic system judges the attributions of a practical state from the changes in its running.

The behavior-based diagnostic system is a dynamic modular system. The number of the diagnostic sub-modules depends on the actual operation of the equipment. Along with the gradually increasing of the fault category, the system can automatically add a number of corresponding new sub-modules to deal with the new fault (Montmain J.& Gentil S.,2000).

The behavior based diagnosis system has another advantage that in the circumstances lacking the prior knowledge of the equipment, it also can establish a effective diagnostic system through the interaction with equipment (Montmain J.& Gentil S.,2000). The mechanism is that: the system uses a monitoring sub-module to recognize the normal behavior. When anomaly occurs, the system will automatically add a new diagnostic sub-module to complete the recognition of the fault thereby increasing the diagnostic capacity and eventually forming a sound intelligent diagnosis system.

It could gradually evolve as a senior intelligent system, and greatly reduce the scale of developing difficulties of the diagnostic system.

Research on Machine Learning Technology

The expert system represents the intelligent systems called "smart reproduction". This design has a purpose of simulating intelligent action for an application reflecting intelligent actions, and it separates the smart performance and the intelligence acquisition. The separation does harm to the flexibility and robustness of the system.

The critical issue of the intelligent system is the learning ability of the system. Automatic knowledge acquisition is always the difficulty in the intelligent research (Jiang Y., Chen X. & Yang X.,2002).

Machine learning is the primary means of improving the intelligence of the system. There is the trend of establishing practical machine learning mechanism. It will make the machine have learning ability to acquire knowledge by itself, modify the knowledge in the database automatically in order to enrich and refine the knowledge in the knowledge base.

SUMMARY

In this chapter, the development and history of the fault diagnosis have been introduced. Then the current widely accepted methods of intelligent diagnosis and their properties are discussed. At the same time, an outlook about the developing trend of the intelligent fault diagnosis is reviewed.

The intelligent diagnostic system is a highly complex systematic project. It is concerned with the disciplines of fault diagnose technology, artificial intelligent technology, pattern recognition technology, computer network technology and communication technology, and there are many problems to be resolved and researched on.

Recently, along with the development of the computer and network technology, the research on Artificial Intelligence has get a more powerful technical platform, and AI technology has been applied into various areas. The model of equipment monitoring and diagnosis system has developed from signal monitoring and diagnosis system to distributed monitoring and diagnosis system (DMDS), and then to Web-based remote monitoring and diagnosis system.

Fault diagnosis system is to, in certain circumstances, identify the causes of the system failure or fault, and judge the state of the failure. Making use of hardware and software systems that simulate brain functions, intelligent diagnosis system does state recognition, diagnosis and prediction on the objectives with the necessary external electronic equipment. Intelligent diagnosis system is inseparable from the hardware and software

systems that simulate brain functions, also it does not exclude the role played by humanity and can achieve human-machine joint diagnosis.

Overall, the development of intelligent diagnosis technology has experienced from single to integration, from concrete to abstract, from local to the network, and from simple to complex.

The modern manufacturing is expected to be flexible integral, intelligent, distributed reconfigurable and global manufacturing. Under the new manufacturing model, in order to meet the requirements proposed by the modern manufacturing systems, monitoring and diagnosis system has the new trend of development. In architecture, it is to be networked and distributed; in the functional structure, it is to be independent and open; in physical structure, it is to be modular and plug-in. In this new trend, internet-based remote monitoring diagnostic system is introduced.

There are many Artificial Intelligence technologies have been applied into the field of intelligent diagnosis. Knowledge-based intelligent fault diagnosis technology is one of the most mature development directions of equipment diagnosis and it is the intelligent diagnostic technology that has been studied and applied the most. However, expert system is facing a major problem, that is the bottleneck problem named knowledge acquisition, the application of machine learning methods in expert systems is expected to solve this difficult problem. While another attempt is intelligent fault diagnose based on neural network, focusing on two aspects. The first one is from the view of pattern recognition, take neural networks as classifiers for fault diagnosis, the other one is from the perspective of prediction, and take neural networks as dynamic forecasting model for failure prediction. But the two each have their own advantages and if they are combined, they can complement each other. There are two ways for the combination, the first one is to establish expert systems by making use of neural networks, and the second one is that the neural network and expert system coexist in a system and constitute a mixed system.

Intelligent diagnosis expert system based on neural network has many advantages that traditional expert systems do not have, but the current neural network expert systems have some weakness, the most fundamental point is that neural network can only simulate intelligent activities on the level of human feeling, so people are trying to study integrated intelligent diagnosis system that combined symbol reasoning and numerical reasoning.

Case-based reasoning can solve new problems by amending the successful outcomes of the similar problems and is a method which solves new problems guiding by the past experience. Because of the uncertainty of some systems' status or the incomplete information, the diagnosis methods based on fuzzy theory are very important.

Remote diagnosis system, by making use of network technology, integrates the resources all around the world and forms a diagnostic system which is extremely powerful in the theory. It also can determine the distribution of the entire system flexibly, solve equipment problems that may arise in the shortest possible time, and reduce the cost to the maximum degree.

In line with the development of the distributed network diagnostic technology, the diagnosis system based on agent technology has good prospects for development. In multi-agent based diagnosis system, the diagnosis tasks are distributed to a number of agents with different functions. The fault diagnosis problems are solved by cooperation and competition of different agents. In macro-respective, multi-agent systems are the social simulation, in micro-perspective, they are different entities with different functions and developed for different use. Communication is essential for the agent cooperation. A number of criteria can be used to implement the agent communication. CORBA established by the organization OMG is the mainly standard for distributed object-oriented application systems. And their properties are very

suitable for distributed, platform-independent communication.

The direct purpose of fault diagnosis is to improve diagnostic speed and precision, so in future development directions of diagnose technology are integrated with fault-tolerant control, redundant control, autonomous control. The future studies can be attributed mainly to the following areas:

1. Multi-sensor data fusion technology.
2. On-line real-time fault detection algorithms and technology.
3. The diagnostic methods of nonlinear dynamic systems.
4. Hybrid intelligent diagnostic techniques.
5. Remote distributed collaboration diagnostic techniques.

With the development of computer technology, signal processing, artificial intelligence and pattern recognition technology, the fault diagnosis technology has developed a lot, but it can not solve all the problems facing the fault diagnosis system. In order to solve the problems, firstly, the traditional diagnostic methods should be studied further and deeper, then constantly enrich the intelligent fault diagnosis methods and improve the fault diagnostic capacity; secondly, the traditional diagnostic methods should be integrated and then make full use of the advantages of various methods; moreover, demand-oriented fault diagnosis method should be promoted to a higher level.

REFERENCES

Alegha, F. (1997). A Remote Controlled Automated Measurement System. IEEE Instrument & Measurment Technology Conferenee, May 19-21, (pp. 1186-1189).

Angeli, C. (2001). Fault Prediction and Compensation Functions in a Diagnostic Knowledge-Based System for Hydraulic Systems. *Journal of Intelligent & Robotic Systems*, *25*(2), 153–165. doi:10.1023/A:1008077902102

Asakura, T. & Kobayashi, T. (2001). Using Neural Networks[J].*JSME International engineering/ Japan society of mechanical engineer. Fault Diagnosis System for Machines Journal. Series B. Fluids and thermal, 43*(2), 364-371.

Breiman, L. (2001). Random Forests [J]. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324

Golobardes, E., Lloro, X., & Salamo, M. (2002). Computer Aided Diagnosis with Case-based Reasoning and Genetic Algorithms. *Knowledge-Based Systems*, *15*(1-2), 42–52. doi:10.1016/ S0950-7051(01)00120-4

Hamscher, W., Console, L., & de Kleer, J. (1992). *Readings in Model-Based Diagnosis*. San Mateo, CA: Morgan Kaufmann.

Han, Y. L. (2004). *Research on Remote Intelligent Diagnosis Technology and its Application Facing to Complicated Equipment*. Shanghai: Shanghai University.

Jiang, Y., Chen, X., & Yang, X. (2002). Intelligent Fault Diagnosis Research and Development. *Information Technology*, *21*(2), 12–15.

Li, M. & Wei, Z.S. (2007). Research of Intelligent Fault Diagnosis System based on Multi-agent. *Fire Control and Command Control, 132*(17).

Liang, X.D. (2000). Intelligent Fault Diagnosis Technology of the Present Situation and Development[J]. *Information and Development,* (3), 24- 23

Liu, G. Q., Chen, X. P., & Fan, Y. (2001). A Formal Model of Multi-agent Cooperative Systems. *Journal of Computer*, *24*(5), 529–535.

Montmain, J., & Gentil, S. (2000). Dynamic Causal Model Diagnostic Reasoning for Online Technical Process Supervision [J]. *Automatica*, *36*(8), 1137–1152. doi:10.1016/S0005-1098(00)00024-8

Noel, S. O'Berry. (2002). Combinatorial Analysis of Network Security. *Proceedings of SPIE. The International Society for Optical Engineering, 4738*, 140-149

Qian, Z.Q. (2004). Equipment fault diagnosis method. *Modern manufacturing engineering, 7*, 105-107.

Raghunathan, R. (2000). A Fast Training Neural Network and its Updating for Incipient Fault Detection and Diagnosis. *Computers & Chemical Engineering*, *24*, 431–437. doi:10.1016/S0098-1354(00)00434-8

Sallans, B. (2002). *Reinforcement Larning for Factored Market Decision Processes*. Ph. D. Thesis. Department of Computer Science, University of Toronto, Canada.

Stanek, M., Morari, M., & Frhlich, K. (2001). Model-aided Diagnosis: an Inexpensive Combination of Model-based and Case-based Condition Assessment [J]. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *31*(2), 137–145. doi:10.1109/5326.941838

Wong, S. V., & Hamouda, A. M. S. (2002). A Logic based Expert System for Machine Ability data-on-demand on the Internet. *Journal of Materials Processing Technology*, *124*(2), 57–66. doi:10.1016/S0924-0136(02)00091-2

Wu, J. P., & Xiao, J. G. (1997). *Intelligent fault diagnosis and expert systems*. Marrickville, Australia: Science Press.

Yager, R. (2001). Penalizing Strategic Preference Manipulation in Multi-agent Decision Making [J]. *IEEE Transactions on Fuzzy Systems*, *9*(3), 393–403. doi:10.1109/91.928736

Yuan, Z. R. (1999). *Artificial Neural Network and its Applications*. Beijing: Qinghua University Press.

Zhang, Q. B. (1997). *The development of diagnostic equipment*. Metallurgical Equipment.

Zhou, Z. D. (2004). *Intelligent control based on network environment*. Defense Industry Press.

Chapter 13
Intelligent Management Information System

INTRODUCTION

Background and Motivation

The Intelligent Management Information System (IMIS) has the potential to transform human decision making by combining research in Artificial Intelligence, Information Technology, and Systems Engineering. The field of Intelligent Decision Making (IDM) is expanding rapidly, due in part to advances in artificial intelligence and network-centric environments that can deliver the technology. Communication and coordination between dispersed systems can deliver just-in-time information, real-time processing, collaborative environments, and globally up-to-date information to the human decision maker. At the same time, artificial intelligence techniques have demonstrated that they have matured sufficiently to provide computational assistance to humans in practical applications. It is the development direction of modern management science and technology. In this chapter, firstly we introduce the introduction and background of IMIS, and briefly, the related design conception. Subsequently, the Intelligent Decision Support System (IDSS) is depicted, which is the most significant technology of IMIS and related activities in the manufacturing process. The applications of IDSS and two cases for industrial manufacturing are then presented, representing the future development direction of manufacturing management. Lastly, a summary of this chapter is given.

IMIS researchers and technologists have built and investigated Decision Support Systems (DSS) for more than 35 years. The developments in DSS began with building model-oriented DSS in the late 1960s which were followed by theory developments in the 1970s, and the implementation of financial planning systems and Group DSS in the early and mid-1980s. During the mid-1980s, Intelligent DSS were implemented through combining knowledge systems with DSS. These developments are discussed below, as well as the origins of Executive Information Systems, On-line Analytical Processing (OLAP), Business Intelligence, and the implementation of Web-based DSS in the mid-1990s, which quickly became a topic for active discussion, and whose influence spread widely.

DOI: 10.4018/978-1-60566-864-2.ch013

The development process of the computer management system can be divided into the following phases.

The first phase (initial stage), 1950s to 1960s: electronic data processing and transaction processing systems, the main tasks being wage management, data statistics, account management, registration statements, and other data-processing services and information services.

The second phase, 1960s to 1970s: the management information system (MIS) and office automation system (OAS). MIS is the development of electronic data processing and is mainly business-oriented management; OAS is the development of business systems, which focuses mainly on office transaction processing and information services.

The third phase, 1980: high-level, strategic and large-scale management decision-making, mainly represented by Decision Support Systems (DSS), using decision analysis methods for solving unstructured and semi-structured information processing problems.

Basic Concepts and Foundations

The Intelligent Management Information System (IMIS) is the next generation of computer management system. It researches ways to improve the intelligence level of computer management systems as well as the design theories, methods and technological achievements of the Intelligent Management System. IMIS is based on the Management Information System and the application of artificial intelligence technology. It is a new computer Management System which has the characteristics of intelligent integration and coordination.

The characteristics of IMIS are as follows:

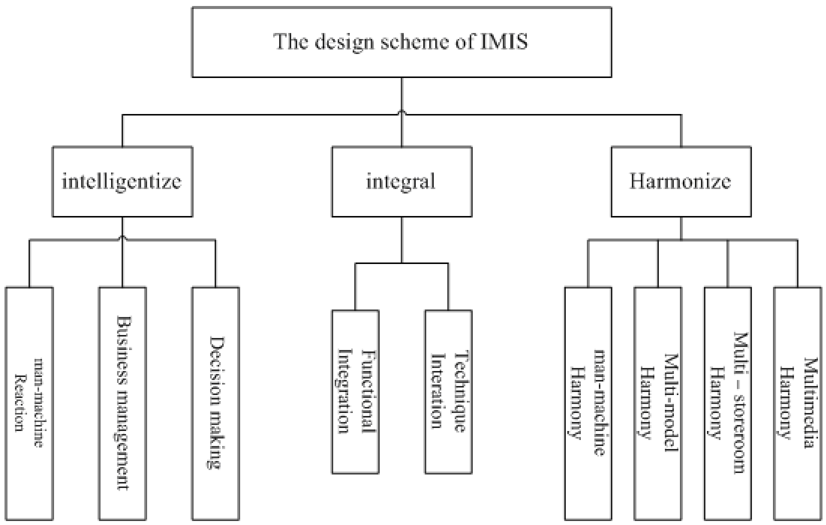
1. IMIS is able to meet the different requirements of various levels of management staff, such as decision-making support for executive staff, planning and scheduling for middle-level managers, and production control and office transaction processing for general personnel.
2. IMIS provides a multi-level, multi-stage and multi-directional information management service:
    a. **Multi-level management:** high-level decision support (macro, large-scale and coarse-grained strategic planning and management); middle-level information management (cope and size of the factory, workshop, department planning and scheduling operation management); activities at base level (micro, small, fine-grained workshop, statistical enquiries).
    b. **Multi-stage management:** short, medium and long-term planning and management; pre-planned management; medium-term management scheduling; post-evaluation of management.
    c. **Multi-directional management:** production management (business administration), financial management, sales management, labor, personnel management, material management and customer management.

IMIS METHODOLOGIES

The Design of IMIS

IMIS is a new generation of computer management system, developed on the basis of the function and technology integration of MIS, OAS and DSS; it applies artificial intelligence expert systems, knowledge engineering, pattern recognition, artificial neural networks, and other methods and techniques. These basic technology have been mentioned exactly in Chapter 2, 4 and 5 of this book. The design idea of IMIS principally includes man-machine harmony, integrated management functions, three-dimensional models and integrated intelligence systems.

Figure 1. The Design scheme of IMIS



IMIS incorporates the design of the main characteristics, which are the plane coordinated, integrated management functions, three-dimensional models of intelligent and integrated systems

The design core of IMIS is to realize the intelligence, integration and coordination of the management information system. Intelligent means that the theory, methods and techniques of artificial intelligence are applied to the development of the management information system for instance, by using neural network technology to explain the portfolio optimization problem, and to develop neural network management information systems; by using pattern recognition and voice technology to develop office multi-media systems; by using expert systems and knowledge engineering technology to develop intelligent decision support systems, and so on.

The so-called 'integrated' includes functional integration and technology integration. Functional integration means that IMIS should have a multi-level, multifaceted, multi-stage all-round management service, and technology integration refers to the integration of model, methods and software technology. Coordination means that the management information system is the realization of the need for a multi-model coordination and other multimedia content. The Intelligent Design Management Information System programmer is shown in Figure 1.

Intelligent Decision Support System

The Intelligent Decision Support System (IDSS) is an application direction of the intelligent management system. IDSS is based on the support system and integrates the Artificial Intelligence Expert System (AIES).

The Decision Support System (DSS) is the comprehensive utilization of the available types of data and information, and the organic combination of many models (mathematical models and data-processing models, etc.) which support decision makers at all levels tin achieving a scientific decision-making system through human-computer interaction. IDSS is a new decision support system based DSS and combined with artificial intelligence technology. It is characterized not only by artificial intelligence technology used in problem-solving in the form of qualitative analysis, but also through comprehensive statistics and operations research to solve the problem of

quantitative analysis, and through qualitative and quantitative analysis of the organic integration, to improve the decision support system applications and the quality of decision-making.

Decision-makers in the field of senior management frequently encountered structural problems for which standard problem-solving procedures could not be used, as these problems could not be accurately described and were highly complex. To solve this problem, the DSS came into being. It has been more than twenty years since the inception of DSS; during that time, it has developed rapidly in terms of the concept, structural design and applied research. Nevertheless, as the emphasis of the traditional DSS was on supporting the integration of data and models, it has some shortcomings, such as an inadequate and unfriendly user-interface. The traditional DSS is characterized by data and math model analysis technology and therefore has many limitations.

Since the 1980s, Artificial Intelligence (AI), and in particular Expert System (ES) technology, has injected new vitality into the vigorous development of the DSS. The issue of how to combine DSS with AI, in particular with the ES, which is a smart Decision Support System, to create the Intelligent Decision Support System (IDSS), also called the Knowledge-Based Decision Support System (KB-DSS), has become an important area for study.

The core idea of IDSS is to combine artificial intelligence technology and other related disciplines and technology, so that DSS has the characteristic of artificial intelligence and can take full advantage of human knowledge.

With the development of IDSS, people have continued to extend the intelligent components of IDSS, which are not only limited to the use of the knowledge base, but also joined with other subsystem components. Model as its part can realize automatic selection and generation of model.

For the human-computer interface, IDSS can make itself easier to use, and can understand the thinking of policy makers by means of its learning function.

For the database, the applications of data warehousing, online analytical processing and data mining technology applications, it can analyze complex data, excavate the implicit knowledge from the data (storage), and enhance the original knowledge base in order to enhance the ability of decision-making.

The Basic Structure of IDSS

By comprehensive analysis of existing IDSS research, it can be concluded that the realization of intelligence can be divided into three elements: the use of AI; the use of the new tools in the database field, such as data warehousing, online analytical processing and knowledge mining technology; the use of other techniques to achieve the reunification of various components of IDSS to make IDSS unitive and coordinative. By using AI to realize the intelligence of the system, IDSS can be divided into three kinds: the traditional IDSS. the IDSS based on machine learning and the IDSS based on the new technology of AI. The structure of IDSS is shown in Figure 2.

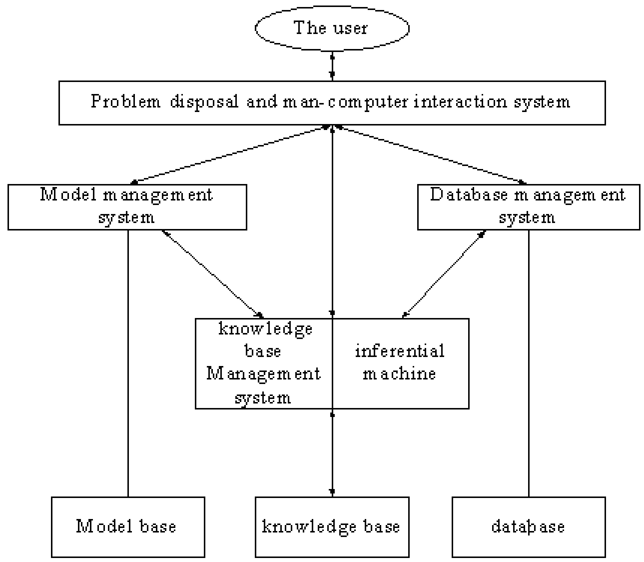
DSS is mainly made up of a dialogue subsystem, database systems and model sub-systems. ES is a more mature application of AI which is mainly made up of the general knowledge base, reasoning and the knowledge base management system. It uses the logic of non-quantifiable statement to express knowledge, and uses an automatic reasoning approach to problem solving. The IDSS is the union of DSS and ES.

The Classification of IDSS

Data-Driven DSS

Data-driven DSS is a type of DSS that emphasizes access to and manipulation of a time-series of internal company data and sometimes external data. Simple file systems accessed by query and

Figure 2. The structure of IDSS



retrieval tools provide the most elementary level of functionality. Data warehouse systems that allow the manipulation of data by computerized tools tailored to a specific task and setting, or by more general tools and operators provide additional functionality. Data-driven DSS with On-line Analytical Processing (OLAP) provides the highest level of functionality and decision support that is linked to the analysis of large collections of historical data. Executive Information Systems (EIS) and Geographic Information Systems (GIS) are special purpose Data-Driven DSS.

Metadata is data which describes the content, quality, condition, and other characteristics of data. It plays an important role not only in the design, implementation and maintenance of the data warehouse, but also in data organizing, information querying and result understanding. It usually records the location and description of warehouse system components. Here, we expand the scope of the metadata, using it to describe and manage the data and environment of the whole system that includes not only the data in the data warehouse platform, but also the task model and algorithms (or functions) in Extract-Transform-Load (ETL) and data mining. Metadata holds the core position of the whole system since it integrates ETL, the data warehouse, and data mining tools. It controls the whole flow from ETL, data warehouse to data mining, so that we can define and execute ETL and data mining tasks more conveniently and effectively.

In MSMiner, the contents of the metadata are as follows:

1. Description of the external data source. The external data source can be a relational database or other kind of data, such as Excel data, plain text, XML text, and so on. In metadata, it contains allocated position and environment information of the external data source, data structure and description of the contents.
2. Descriptions of the subject, including the name and remark of the subject, when the subject is created and updated, etc.
3. Description of databases under a subject, including the name, type and remark of database, the login information and other information.

4. Description of tables in a database, including fact tables, dimensional tables and temporary tables. It contains some table information and field information.
5. Description of the ETL task, containing organization and steps of the task, data source, selection of the transformation functions, assumption of the parameters, creation and execution history of the task, and so on.
6. Description of the data mining task, containing organization and steps of the task, data source, selection of the mining algorithms, assumption of the parameters, evaluation and output of the results, creation and execution history of the task, and so on.
7. Description of the data cube, containing dimension and measure of extracted information, building information of the star-structure, and so on.
8. Management of the algorithm base for data mining, containing the registration and management of the mining algorithms.
9. Management of the functions for ETL, containing the registration and management of the functions.
10. User information, containing the user's basic information, authority, operational history, and so on.

We build the corresponding metadata classes using the object-oriented method. We take the three-tier architecture as the system architecture and put the metadata management subsystem at the middle tier position, where it can be regarded as a metadata management server. The upper tier accesses and manages metadata via the middle tier.

Metadata is automatically generated while every component of the system is created. Metadata will be changed during the daily maintenance of the system. MSMiner provides a special metadata manager subsystem that can maintain the metadata directly and ensure that the whole system is managed validly.

ETL is an important subsystem of MSMiner. The main purpose of the ETL function module is to transform the operational data from the source database to analytical data in the data warehouse. The data in the data warehouse is integrated and extracted from diverse databases (for example Oracle, SQL Server, Access, Foxpro, Excel, DB2), and there are many differences between the operational data in the source database and the analytical data in the data warehouse. Loading the data from various data sources directly into the data warehouse is not good practice. To obtain clean data for data warehouse, the data from the previous database must be cleaned, collected and transformed before being integrated into the data warehouse. It is a key and complex step during the building of the data warehouse. Generally speaking, the ETL subsystem needs complete the following actions:

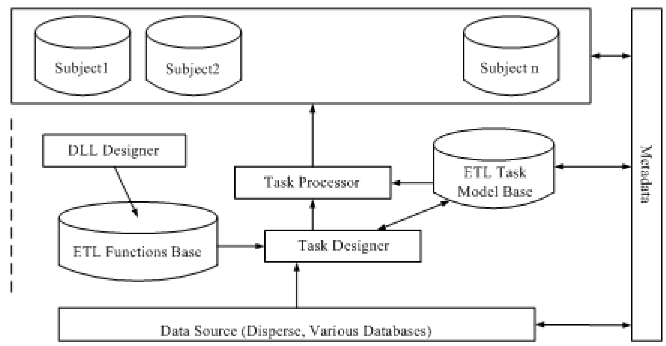
1. Because of data repetition and conflict in the source data from disperse database, the subsystem should unify the conflict data.
2. To get the comprehensive data in the data warehouse, the subsystem should transform the original data structure from an application-oriented one to a subject-oriented one and do some generating and computing.

The basic architecture of the ETL subsystem is shown in Figure 3.

From this figure, it is clear that there are 4 modules in the ETL subsystem:

1. **The friendly user interface:** Users can conduct ETL operations expediently using this interface, such as designing ETL tasks, registering new ETL DLL functions, scheduling and executing ETL tasks, and examining the result of ETL tasks.
2. **The integrated ETL functions management and ETL task management:** This module includes registering new ETL DLL functions, building new ETL tasks,

Figure 3. The basic architecture of ETL sub-system



scheduling and processing of ETL tasks and suchlike.

3. **The uniform metadata management:** The whole subsystem is developed in a metadata-oriented way, which is to say that all information of this subsystem, including data source, algorithm and result, is managed by metadata.
4. **The database server:** The ETL subsystem supports diverse databases, such as Oracle, SQL Server, Access, Foxpro, Excel, and DB2.

The subsystem supports the expandable ETL function base. The main algorithms for the ETL function are realized in the form of dynamic link lib (DLL) with uniform interfaces. Users can design the ETL task according to their need by choosing the relevant ETL DLLs. At present the subsystem provides approximately 30 kinds of ETL DLLs. In addition, users can develop new ETL DLLs in accordance with uniform interfaces and add them into the ETL function base. In order to improve efficiency, the ETL tasks can be scheduled at a designated time and processed concurrently.

The data warehouse is ‘a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decision’. The function of the data warehouse is to provide a general data warehouse environment by which users can create and maintain their data warehouse in accordance with different needs to complete data analysis and processing and preparate for the data mining task.

The data warehouse in MSMiner consists of many subjects. When a data warehouse is created, users establish several subject fields according to the application needs, and the system helps users to extract the data for each subject and to model them by star-schema. Based on the above operation, the data warehouse realizes the multi-dimension data cube and OLAP, and provides a valid data source for data mining and decision-making. The final results may be shown by visualization tools.

The data warehouse in MSMiner is modeled by star-schema. The system extracts the data from source tables or views and builds multiple fact tables through data extraction, transformation and loading by the subject’s request. A star-model’s structure is made of one fact table and several dimension tables related to the fact table, where the fact table includes multiple dimensions and measures. The dimension stands for the special visual angle for viewing data, such as time dimension, distribution dimension, product dimension and so on. The measure is the data’s real meaning and describes what the data is. Each dimension table describes a certain dimension and its values, and each dimension consists of several levels. For example, a time dimension may be divided into three levels: year, season, and month, as each describes a different query layer. One or several

star-schema structures form a subject, which is the basic unit of the data warehouse.

The OLAP is realized in two ways: by creating a special multi-dimension database system (MOLAP) and by simulating the multi-dimension data by using the relational database (ROLAP). MSMiner supports ROLAP, which is based on the star schema. The star structure related to the multiple dimension tables simulates the multi-dimension data cube, where the dimensions and measures in the data cube come from dimensions and measures in the star schema. When OLAP operations are executed in the data cube, multi-dimension analysis translates the request into SQL statements and queries in fact tables, then shows the results in the form of multi-dimension data.

At present the system supports standard OLAP operations, such as slice, dice, roll up, drill down and pivot. The results may be displayed in many forms such as cross-tabulation tables, bar charts, pie charts or other forms of graphical output.

The results of OLAP operations and data in fact tables may be the data source for the data mining subsystem. They may be helpful in preparation work for data mining.

Model-Driven DSS

Model-Driven DSS emphasize access to and manipulation of a model, for example, statistical, financial, optimization and/or simulation models. Simple statistical and analytical tools provide the most elementary level of functionality. Some OLAP systems that allow complex analysis of data may be classified as hybrid DSS systems providing both modeling and data retrieval and data summarization functionality. In general, model-driven DSS use complex financial, simulation, optimization or multi-criteria models to provide decision support. Model-driven DSS use data and parameters provided by decision makers to aid decision makers in analyzing a situation, but they are not usually data intensive; that is, very large databases are usually not need for model-driven DSS. Early versions of Model-Driven DSS were called Computationally Oriented DSS by Bonczek, Holsapple and Whinston (1981). Such systems have also been called model-oriented or model-based decision support systems.

Knowledge-Driven DSS

Knowledge-Driven DSS can suggest or recommend actions to managers. These DSS are person-computer systems with specialized problem-solving expertise. The “expertise” consists of knowledge about a particular domain, understanding of problems within that domain, and “skill” at solving some of these problems. A related concept is Data Mining. It refers to a class of analytical applications that search for hidden patterns in a database. Data mining is the process of sifting through large amounts of data to produce data content relationships. Tools used for building Knowledge-Driven DSS are sometimes called Intelligent Decision Support methods (Zhongzhi Shi, 1988, Dhar and Stein, 1997).

Web-Based DSS

Web-Based DSS deliver decision support information or decision support tools to a manager or business analyst using a “thin-client” Web browser like Netscape Navigator or Internet Explorer that is accessing the Global Internet or a corporate intranet. The computer server that is hosting the DSS application is linked to the user’s computer by a network with the TCP/IP protocol. Web-Based DSS can be communications-driven, data-driven, document-driven, knowledge-driven, model-driven or hybrid. Web technologies can be used to implement any category or type of DSS. Web-based means the entire application is implemented using Web technologies; Web-enabled means that key parts of an application such as a database remain on a legacy system, but the

application can be accessed from a Web-based component and displayed in a browser.

Simulation-Based DSS

Simulation-Based DSS deliver decision support information or decision support tools to help managers analyze semi-structured problems through simulation. These diverse systems were previously all called Decision Support Systems. DSS could support operations, financial management and strategic decision-making. A variety of models were used in DSS including optimization and simulation.

GIS-Based DSS

GIS-Based DSS deliver decision support information or decision support tools to a manager or business analyst using GIS. General-purpose GIS tools are programs such as ARC/INFO, MAPInfo and ArcView that have extensive functionality and can be difficult for users unfamiliar with GIS and cartographic principles to learn. Specific-purpose GIS tools are programs that are written by a GIS programmer to provide a user group with specific functions in an easy-to-use package. In the past, specific-purpose GIS tools were written primarily using a macro language. This method of delivering specific-purpose GIS tools requires that each user have a copy of the host program (ARC/INFO or ArcView) to run the macro language application. The GIS programmers now have a far richer set of tools for application development. Programming libraries with classes for interactive mapping and spatial analysis functions have made it possible to develop specific-purpose GIS tools using industry-standard programming languages that can be compiled and run without a host program (stand-alone). Internet development tools have also matured, making it possible to develop fairly complex GIS-based programs that users can use through the World Wide Web.

Communication-Driven DSS

Communications-Driven DSS is a type of DSS that emphasizes communications, collaboration and shared decision-making support. A simple bulletin board or threaded email is the most elementary level of functionality. The comp.groupware FAQ defines groupware as “software and hardware for shared interactive environments” intended to support and augment group activity. Groupware is a subset of a broader concept called Collaborative Computing. Communications-Driven DSS enable two or more people to communicate with each other, share information and co-ordinate their activities. Group Decision Support Systems or GDSS is a hybrid type of DSS that allows multiple users to work collaboratively in group work using various software tools. Examples of group support tools are audio conferencing, bulletin boards and web-conferencing, document sharing, electronic mail, computer supported face-to-face meeting software, and interactive video.

CASE STUDY I: MULTI-AGENT IDSS BASED ON BLACKBOARD

Introduction

In current society, the vast majority of work is completed by the collaboration of groups. The decision-making process goes with the completion of most work. To a large extent, the pros and cons of the decision-making process affect the smooth completion of tasks. MAS technology and network technology provide a new way for the establishment of IDSS, and Multi-Agent technology helps to realize an IDSS which is intelligent and integrated. MAS is one of the basic technology of this case, meanwhile, it is also one of the important technology that discussed in the book, and the definition of relative concept refers to chapter 3 of the book.

The blackboard model is a highly structured problem solving model for "timely" problem solving; as an efficient approach of multi-source of knowledge, it is widely used for large knowledge base systems, and it is a common pattern for the establishment of knowledge processing systems. In fact, every knowledge source is equivalent to a particular issue, and many experts will adopt a cooperate and interactive manner when solving an issue. The blackboard system can be used as a Multi-Agent System.

Based on the Blackboard system and Multi-Agent technology, the Multi-Agent intelligent decision support system based on a blackboard was put forward.

In this system, the blackboard provides a public work area for a distributed architecture system and each Agent can use the blackboard to exchange information, data and knowledge; they can visit the blackboard at any time to extract their respective work information needed to complete their task, and by using the blackboard, Agents can coordinate between each other and work together to solve a given problem, integrating the wisdom of decision makers and the system.

Structure

In the Multi-Agent intelligent decision support system (MA-IDSS) based on Blackboard, each decision-making task or function may be completed by an independent Agent, and each Agent analyzes the problem according to their point of view. Each Agent has a different capacity, opinions and beliefs, as well as different problem solving skills. Different Agents examine the decision-making problem from their own perspective and complete certain subtasks independently, and then achieve a common goal through the blackboard collaboration. As Agents are autonomous entities, they can join or leave a problem solving system at any time so that the decision-makers participate freely in the decision-making process, ensuring the flexibility of the system.

Figure 4 shows the three-tier architecture of interactive layer, intelligent decision-making layer and resource layer. The interactive layer is composed of the intelligent interface Agent and the decision-making user. The intelligent decision-making layer is composed of the blackboard, the functions Agent and decision-making Agent. The resource layer is composed of the model warehouse and model library management system, the repository and repository management systems, the methods warehouse and methods warehouse management systems, and data warehouses.

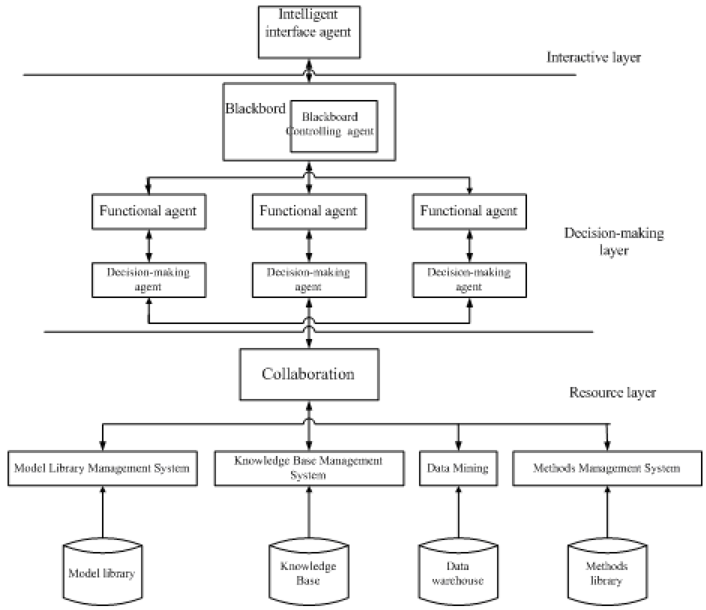
Decision-Making

The decision-making step of MA-IDSS based on Blackboard can be described as: problem input; decomposition of the decision-making problem; solving the problem by collaboration of decision-making Agents; integration of the decision results; forming an overall solution, which is depicted by Daniel.

The steps in detail are as follows:

1. Input information of the relative decision-making problems by the interaction between the decision-maker and the intelligent interface Agent;
2. With the participation of the decision-maker, the blackboard control Agent analyzes all problems, searches the knowledge base, decomposes the decision-making problems into small parts, and assigns them to the corresponding data plane;
3. The function Agent supervises each plane of the blackboard and activates the corresponding decision-making Agent by matching the qualifications and a series of control strategies;
4. Each decision-making Agent uses their own decision-making knowledge to complete their mission, gets the result of problems and changes the assumptions on the blackboard;

Figure 4. The structure of multi-agent IDSS based on blackboard



5. Continuing to interference decision-making, a number of consultations and conflict resolutions take place between the decision-making Agents, under the control of the blackboard Agent;
6. Daniel thought that blackboard controls agent and integrates the sub-decision results of each decision-making Agent to form completed decision schemes. In the decision making process, policy makers have the right to make choices, which eventually achieves satisfactory results.

CASE STUDY II: INTELLIGENT RECONFIGURABLE ERP SYSTEM

Basic Concepts and Foundations

Enterprise Resources Planning (ERP) comprises three levels: management thinking, software product and management system.

1. The well-known U.S. computer technical consulting and assessment group, Garter Group Inc., put forward a set of enterprise management system standards, of which the essence is management thinking of the supply chain, developed on the basis of Manufacturing Resources Planning (MRPII).
2. It is the comprehensive application of the client/server system, relational database structure, object-oriented technology, graphical user interface, the fourth-generation language (4 GL), network communications and other information industry results, and it is a software product which views ERP management thought as its soul.
3. It is an ERP system which integrates enterprise management concepts, business processes, basic data, human resources, computer hardware and software.

Initially, ERP is a management idea based on the enterprise of the supply chain, which expanded the scope of management on the basis of MRPII

and created a new structure. Its basic idea is in viewing enterprise business processes as being closely linked with the supply chain and dividing enterprises into several cooperative subsystems, such as finance, marketing, manufacturing, quality control, maintenance services, engineering technology and so on. This management style was first used by the manufacturing industry and mainly considered the inventory of materials management; as a result, the Material Requirements Planning system (MRP) came into being. At the same time, other departments established their own information management systems, such as the accounting department's computer accounting system and the personnel department's personnel records management system, which were mutually independent initially and lacked correlation. They were "information islands" that did not develop an IT function, but instead caused repetition and discordance in business management. MRPII emerged, in this environment, focusing on "manufacturing and selling the right product at the right time", and managing the "people", "wealth" and "matter" intensively.

ERP can be said to be an extension of MRPII. Firstly, it has shifted the core of system management from "manufacturing and selling the right product at the right time" to "gaining the best increment at the best time and best places". Secondly, based on the transfer of the management core, its management range and the scope have been extended from manufacturing to other industries and enterprises. Thirdly, it has greatly enhanced functional and operational integration in business, and in particular, the introduction of business intelligence has changed the simple affair disposing system into the intelligent management control system.

Comparison

The differency of modern industrial enterprise, except the scale difference, the difference of manufacturing technology and management technology is more important, and the most important is the management difference. As a whole, the following management problems are normaly exact:

1. Production scale is less than normal, which has not yet beccome large-scale production.
2. Marketing information is ineffective, production predicted data is incorrect.
3. The term for designing new product is long which can not adapt to the requirement of fast-changing market.
4. In the production process, the raw material, semifinished product and finished goods takes up massive liqudity because of large stock.
5. The communication among each department of enterprise is not good, long-term, low efficiency and high inaccurancy.
6. The communication and quality control with supplier of raw material and parts does not work effectlvely which effect the delivery time and quality of product..
7. Cost accouting is not detail, most lack cost accounting of parts which can not control cost effectively.
8. Although established the marketing and after sales service network around the country, the the whole network is inefficiency and the information feedback is untimely, which lack modern information technology to support.
9. The advanced management thought based customer-centered is not filtering into people's mind.

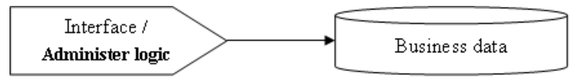
The aforementioned pushes each enterprise to find out effective way in management. That is, EPR system is exactly the result from this kind of requirement. ERP system also changes with the development of enterprise management requirement. Take automobile enterprise management as example, in the beginning of automobile development, the type of organization

of automobile production is make to stock., and the automobile variety is quite single. Such as Ford only product black T in 1920s, and the key of automobile enterprise management in this time is how to improve the product quality, enlarge the production capacity and enhance productivity. Nowadays, with the improvement of people's living standards, people's consumer demands to automobile pay more attention to personality. Automobile production is totally changed from the product-centered in past into customer-centered, and the organization type of automobile enterprise is also changed from build to forecast into build to forecast with make to order together., even some advanced enterprise is totally make to order. For modern automobile enterprise, the management key point is that responding to the order demand of customer quickly, exact and reasonable production schedule and accurate cost control. Detailed, the ERP management requirement of automobile enterprise mainly includes following aspects:

1. Procurement, inventory, production and finance should be more closely with each other.
2. Procurement Plan Management
3. Inventory management is one important part of manufacturing enterprise.
4. Production plan part is the core of manufacturing enterprise, and it is also the most complicated and the hardest part to management of the whole enterprise management.
5. Marketing Management
6. Cost accounting and control is one of the core parts of manufacturing enterprise management.

The earliest ERP software development model integrated interface functions and business logic functions, as shown in Figure 5. This development model did not have any reconstruction characteristic. With the development of software technology, the use of this model of software architecture has become limited nevertheless; this model also has

Figure 5. The early two-tier development model



its merits; for certain pinafore developing a restructurable system if the scale of the application system is not too large and time is not limited. It also has obvious shortcomings, however, and may face a mass of modifications when the system is carried into execution. Interface changes may increase workload, and some functional modules may need to be rewritten; finally, even when the system is up and running, the whole system may appear to suffer confusion, poor stability and poor expansibility.

In the development and implement process of ERP system, there always are two big problems: one is designing the interface between CAD and ERP, another is the addition of production department. These two problems are ERP system to be resolved. In addition, how to design and realize aided decision making of ERP system better so as to realize automation of enterprise management is also a problem. Currently, the espoused development model is that which separates interface functions and business logic functions, as shown in Figure 6. Changes to the general interface of this model will not affect the actuating logic, nor will changes to the performing logic affect interface changes; its flexibility is therefore stronger than the first model, which can reduce the cascading effect. This model is based on the general architecture of components, such as Common Object Request Broker Architecture (CORBA), Distributed Component Object Mode (DCOM), Enterprise Java Beans (EJB) and so on. Through the application of the components, the interface and specific business logic will be separated, and different components will have different functions. Well-designed components can be reused in different systems and modules, thus accelerating

Figure 6. The traditional three-tier system architecture

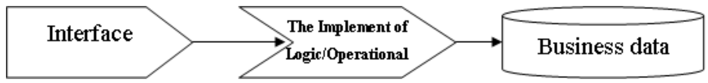
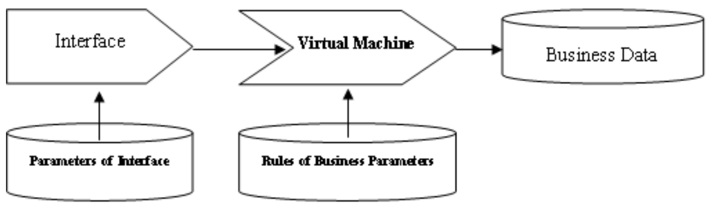


Figure 7. The parametric structure of system



system development speed and enhancing system security and stability. Generally, the connection of these components can only be in accordance with the existing management business process design system. During the concrete implementation and secondary development of the system, a lot of work is still likely to be faced in compiling the code. Therefore, component-based architecture just provides us with a new system design tool or way of thinking; relying only on the design of this system cannot improve the reconstruction of the system.

From the analysis above, we can see that no matter what kind of model, no matter what the realization of technology is, the system implementation process will face the same problem, that is, the code has to be changed constantly, or possibly cannot be modified. This results in a long cycle of system implementation – possibly a year, or even two or three years or longer, and its effectiveness is greatly reduced.

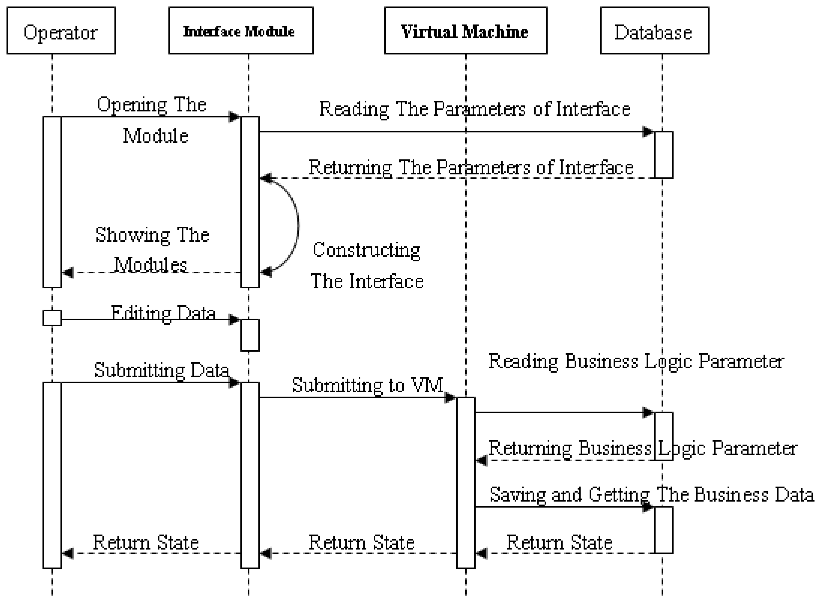
Based on the analysis above and the parameters of the virtual machine of ERP system architecture, the model is entirely achieved through the configuration of parameters from the interface settings to the design of business logic, and only needs modification to the system configuration parameters even the structure changes of the data tables. The system of the hierarchical structure is shown in Figure 7.

Firstly, the system has to realize the separation of the interface function and business logic function, and then to realize the separation of the business logic function and business process. The main characteristic of the system is the realization of implementation logic. The realization of implementation logic is achieved by the cooperation of the virtual machine and virtual machine instructions. The set of virtual machine instructions is realized by setting the parameters; therefore, different business logic can be realized by configuring different business parameters. Business logic changes can be reflected in the system, through the change of operational parameters.

The system interface is realized by reading the dynamic structure of interface configuration parameters; interface changes can therefore be quickly realized through the revision of the interface configuration parameters.

The operation process of the system consists of reading the interface system and setting the parameters, constructing an operation interface dynamically, collecting or editing the data with operation interface, and then submitting the data to the virtual machine. The virtual machine reads the rules of the business parameters, processing

Figure 8. The sequence chart of system working principle



the data. The work process chart is shown in Figure 8.

It can be seen from Figure 8 that the system can be divided into a three-tier structure of logic functions: the presentation layer, affair logic layer, and data layer. Only the data layer is similar to the traditional distributed system structure. The presentation layer is composed of the main program and management controller, the affair logic layer is composed of the virtual machine, and the data layer is composed of the database management systems, data storage systems and business data.

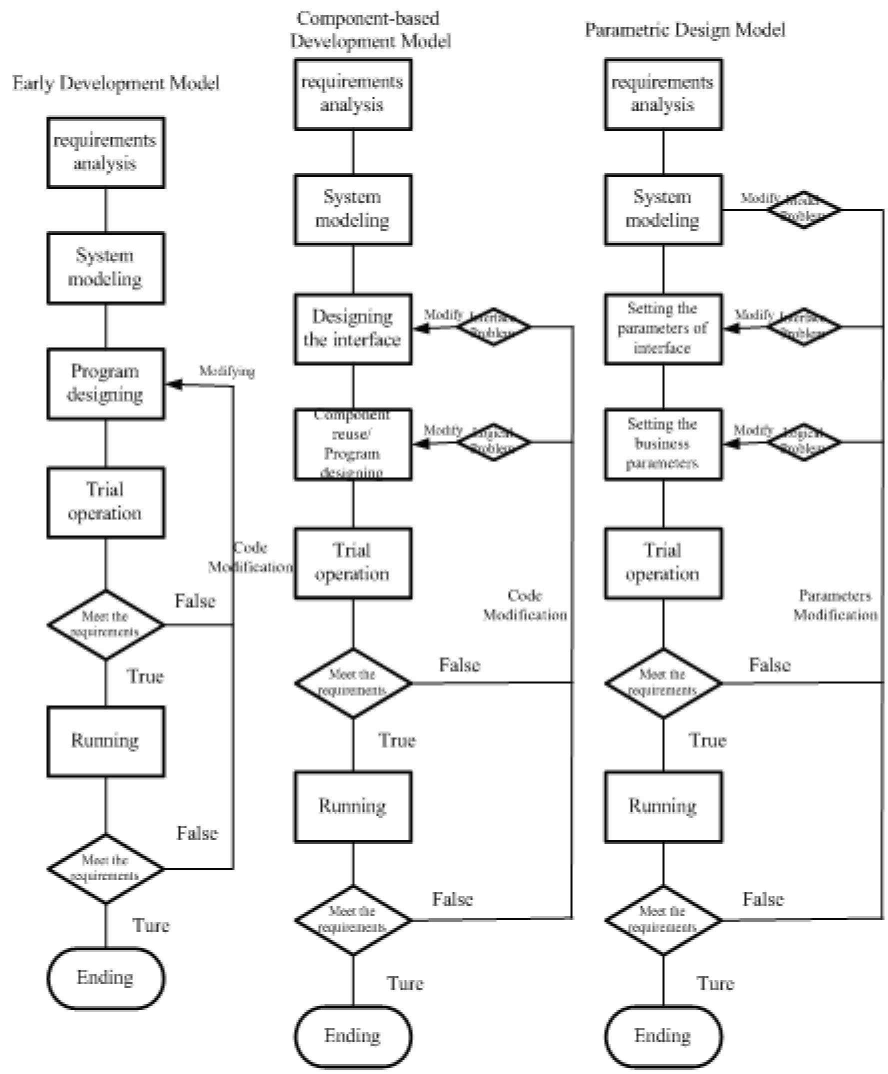
The system contains three parts in the data relationship: the implementation procedures, system data and business data. The implementation procedures and system data constitute the core of the system and the implementation of procedures comes under the system data operation and implementing commercial operations. System data is the description of the commercial data structure and system environment. Business data is the data in the process of enterprises operation.

The system is characterized by lower-level abstraction and a higher degree of abstraction. Not only does it achieve the separation of the interface and the logic of the implementation, but it also realizes the separation of the implementation logic and specific business rules, which realizes the completed parameterized design of the ERP system. Therefore, system development speed is greatly accelerated and flexibility is also enhanced.

The comparison of using the ERP system with this model with the traditional design of ERP system is shown in Figure 9. There are three parts in this figure:

- **Early Development Model:** Interface and logic is combined, the code reusability is poor, a great deal of work needs to be amended, and interface modification may be involved in the logic changes. Once the datasheet is confirmed, it cannot be easily changed; with heavy workloads and long cycles, the system cannot be re-designed once the initial system modeling has been completed.

Figure 9. The comparison of parametric design and traditional design



- **Components-base Development Model:** The system development workload can be reduced by the separation of interface and logic and the reusability of components, but business logic still needs code programming, datasheets cannot be easily changed once the workload is determined, trial operation or changes of business require amendments to the code, and the workload remains large and of long cycle. The system cannot be re-designed once the initial system modeling has been completed.
- **Parametric Design Model:** The interface is separated from the logic, and the business rules and procedures are also separated; the system is designed with a fully parametric model, and the system interface design and business logic design are set parameters; changes of business or interface also need parameter modification; therefore, the system design workload is greatly reduced while its flexibility is enhanced. The structure of the datasheet can be changed at random, and the system can even be re-modeled and re-designed.

Connotation of IR ERP

The Intelligent Reconfigurable ERP (IR ERP) system will not only aim to enhance work efficiency by implementing ERP, but also to establish a comprehensive set of intelligent decision-making systems, which will transfer the focus of business management from transaction to high-level intelligent decision-making systems, such as management, control, planning and analysis of enterprise. The restructural thought is the central to IR ERP. Using parametric reconfigurable technology enhances the understandability, maintainability and reusability of software. The system functional module is shown in Figure 10.

From Figure 10 it can be seen that the system is mainly composed of two parts, the background maintenance tools and the frontend client. The background maintenance tools include form management, view management, menu management, authority management and report management; the frontend client includes the main program, resource management, virtual machine, view examination, reporting output and Byte Order Mark (BOM) subsystem. System data includes two major parts: system parameters and commercial data. System parameters contain various system operating parameters, and business data contains business data of commercial users.

The relationship among various modules is: background maintenance tools are responsible for the maintenance of the system parameters; the system parameter is the basis of the operation of the frontend client; the frontend client operates business data on the basis of system parameters.

The specific functions of each module are as follows.

Forms Management

The Forms Management system is the main maintenance module of the system parameter. The module is responsible for maintaining the storage sheet of commercial data, setting the storage structure of the storage sheet, and the specific type and size of a detailed storage field. It sets the association between the current storage table and other storage tables in the system, and also sets the display and edit mode of the storage sheet in the System Explorer. Setting the parameters in Forms Management is the basis of the explorer operation.

View Management

The System Explorer has the function of seeking information from the resources directly, but its inquiry is limited to a single resource. Setting of the view manager is necessary for inquiring a number of joint resources and drawing complex report forms. The setting of the view manager is similar to the form manager, which can set all kinds of resources as well as the display structure of its inquiry results.

Menu Management

The menu manager is responsible for designing the menu structure in the main program, displaying the title, and registering the new functional module to the system.

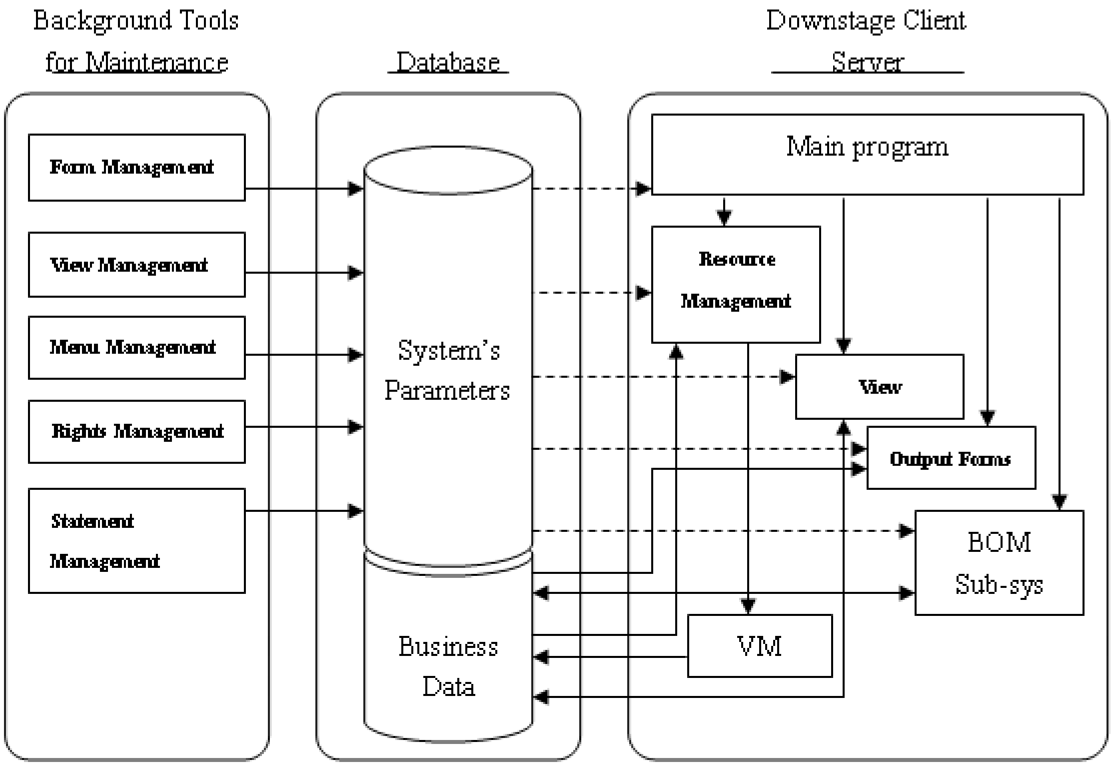
Authority Management

The authority manager is responsible for managing the user permissions. Users can set up sub-designated administrators and ordinary users; the system administrator has access to all functions, whereas ordinary users have only assigned functions which are distributed by the manager.

Statements Management

The statements manager is responsible for the design of the print format, and for using the print format which has been designed to print in Explorer and view manager after registering report forms to the system.

Figure 10. The system functional module



Main Program

The main program is the main operating platform of the system; when a user accesses the system, the first operation of procedures it the main program. The main program is responsible for calling the System Explorer, view manager and other functional modules.

Explorer Management

The System Explorer is responsible for all kinds of resources management, including increase, modify, and delete operations, and can also produce various statistics on current resources.

View Show

View show is responsible for inquiring set view; it can also operate the current statistical view. The setting of view is carried in view manager.

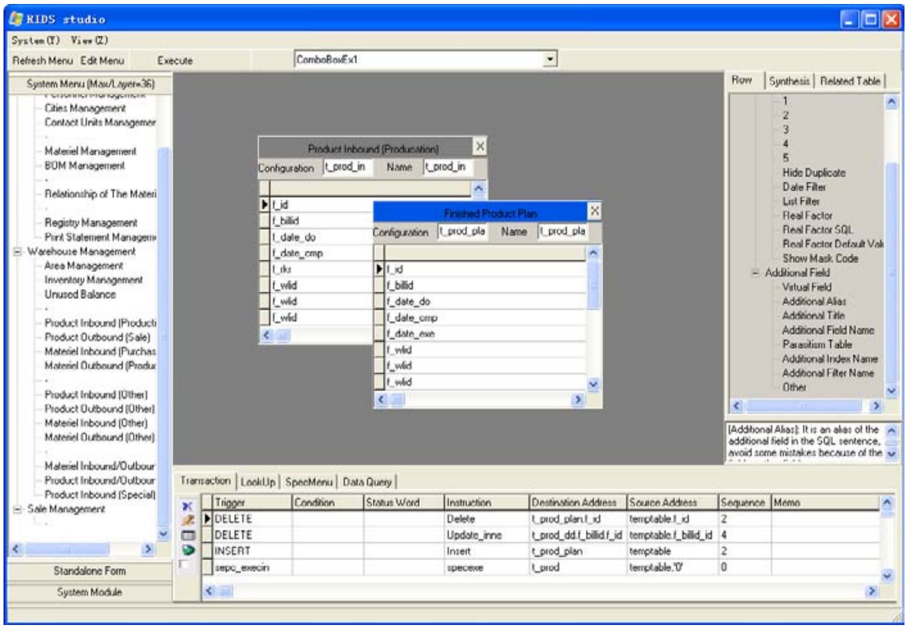
Statements Output

The statements output module is called by the System Explorer and view manager, according to current data and relevant parameters set by statements management, and the statements output module prints the statements.

BOM Subsystem

The BOM subsystem is responsible for the maintenance of product structure data. Since the storage and display of BOM is a logical structure of a tree, its storage and display are more

Figure 11. The system console



troublesome, so it is designed as an antithetical module. Other business data is generally data of two-dimensional form and is maintained by the System Explorer.

Virtual Machine

The virtual machine is mainly called by Explorer. Explorer submits data that is added, edited or deleted to the virtual machine; the virtual machine then processes the data and stores it in the database.

The Primary Console and View Manager of IR ERP

System Primary Console of IR ERP

The interface of the system main console is shown in Figure 11.

The system console includes the menu area, form area, attributes column and virtual machine instruction areas. All levels of menu are displayed in the system menu. Editing the parameters is achieved by clicking on the "Edit menu" button in the top of the menu area. The editorial column that shows the menu parameters is shown in Figure 12.

Each menu item has five parameters: menu headings, called modules, interface function, called configuration and visibility. Menu heading parameters show the content of menu items that appear in the interface; called modules show the name of a specific dynamic link library; if a new module is developed, it is registered in the "system modules" column, as shown in Figure 13; the interface function denotes the output interface parameters of the dynamic link library; the called configuration denotes the configuration names of the module operation, and each configuration stores related parameters of specific forms; visibility denotes whether the menu item is shown in the main interface of the system operation platform. The default state is to tick the check box when adding a menu item, and to cancel the tick if hiding a menu item is required.

After double-clicking a menu item, the corresponding configuration column will be shown in

Figure 12. The editing column of menu items

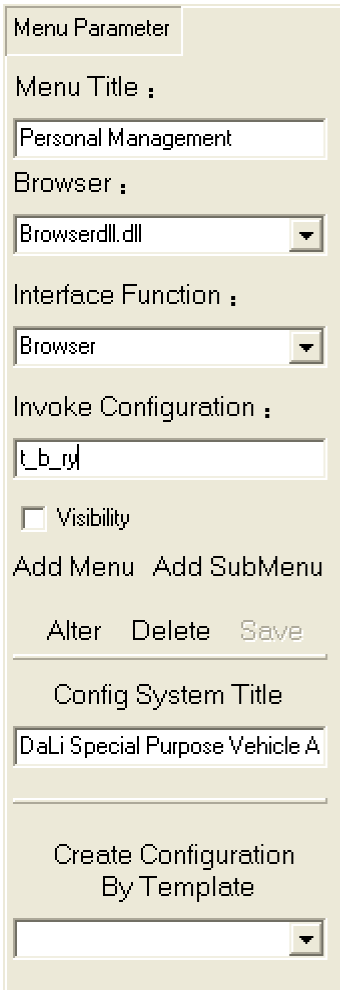
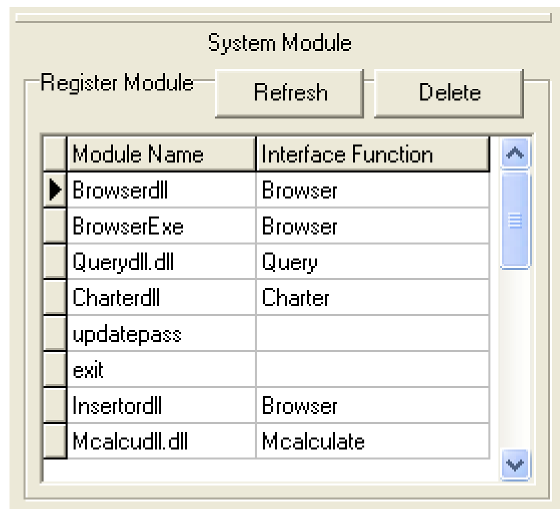


Figure 13. The registration column of system module



the table sheet area, as shown in Figure 15. When an item in the configuration column is clicked, the detailed parameters of this item will be shown in the right side of the property column, as shown in Figure 14. The business rule parameters of the configuration will be shown in the virtual machine instruction area at the bottom after selecting a configuration column, as shown in Figure 16.

The View Manager of IR ERP

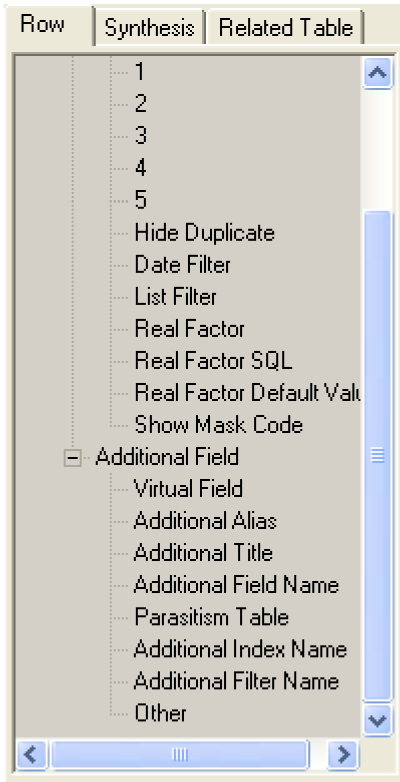
The main interface of the view manager is shown in Figure 17, which includes a view column and a view structure column. The view column shows views which have been established; the main parameters are the view name, SQL statements, pre-treatment and post-processing; the view structure column shows the interface structure of view and parameters of each item.

The Characteristics of IR ERP

The characteristics of the intelligent reconfigurable ERP system are as follows:

1. The system has strong flexibility and high flexibility. It can add needed functions or modify existing functions flexibly with the configuration of parameters.
2. Similar to the browser framework, it is convenient to install and maintain. The upgrade client of the system can be reflected directly once the client is installed without re-installing the client.
3. The system is easy to operate. The unified system interface and simple interface style make use of the system very easy, and users can operate the system after simple training as long as they have basic computer operating knowledge.

Figure 14. The attribute column



4. The system has a powerful function of rights management. It can set each individual's operational authority to ensure data security according to the different duties of operators.
5. It has a flexible function for statistics, which can add new functional modules of statistics and inquiry according to the user's needs at any time.
6. It has an exoteric design function for statements for both statistics and inquiries, and can conveniently output and print in various statement formats.

Example I: IR ERP Systems for Enterprises of Automobile Fitting

Main function modules of the system include.

Basic Data

Department management, personnel management, urban management, related unit management, materials management, management of metal

Figure 15. The configuration of table sheet

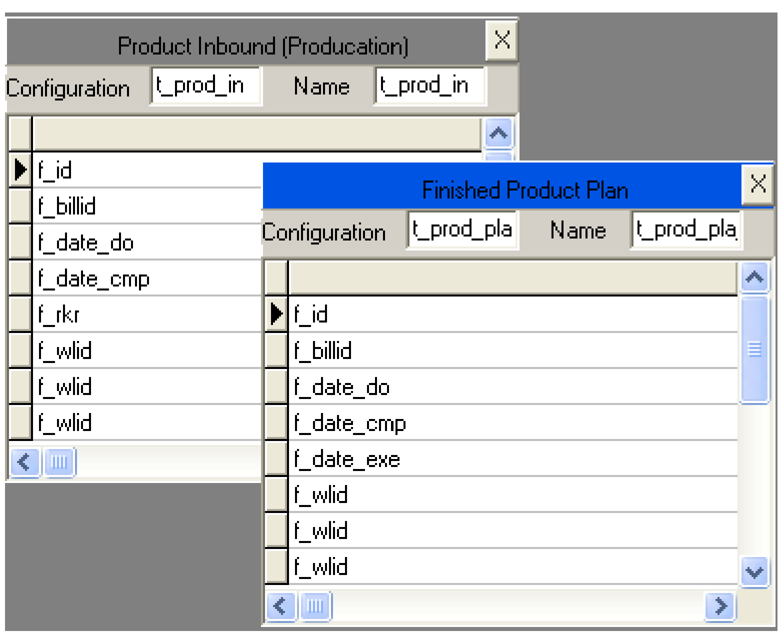
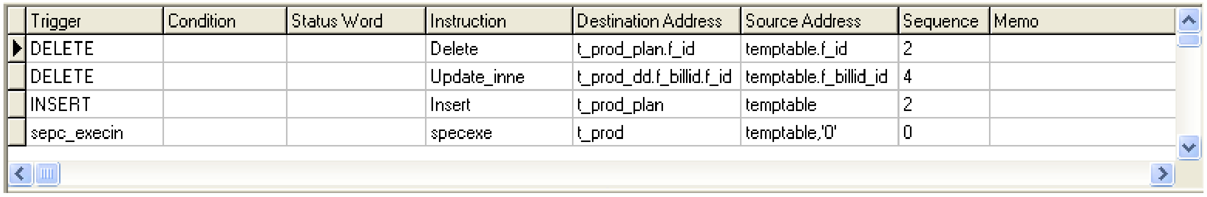
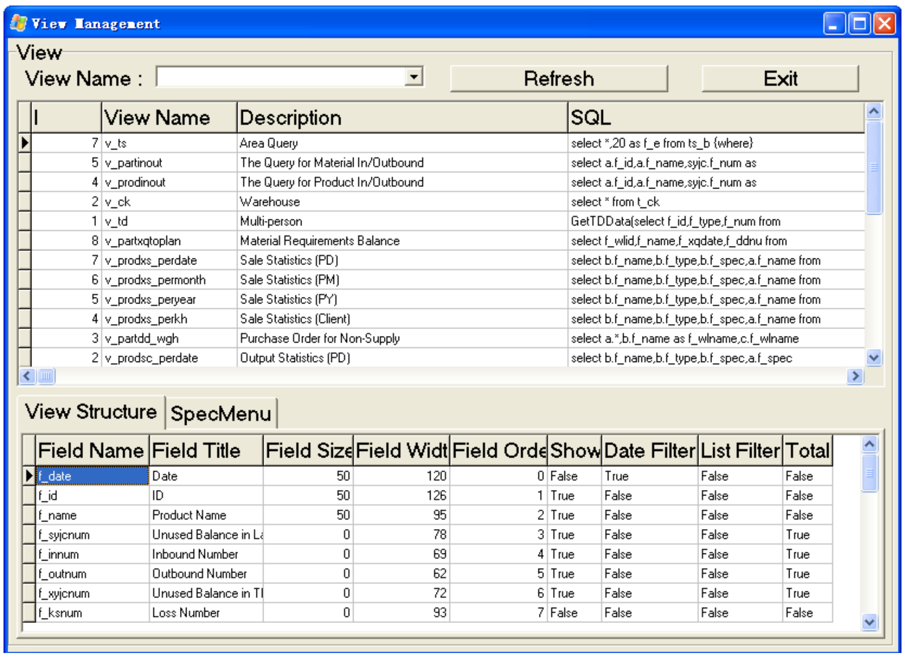


Figure 16. The instruction parameter column of virtual



Trigger	Condition	Status Word	Instruction	Destination Address	Source Address	Sequence	Memo
DELETE			Delete	t_prod_plan.f_id	temptable.f_id	2	
DELETE			Update_inne	t_prod_dd.f_billid.f_id	temptable.f_billid_id	4	
INSERT			Insert	t_prod_plan	temptable	2	
sepc_execin			specexe	t_prod	temptable,'0'	0	

Figure 17. The view manager



View Management

View

View Name :

Refresh

Exit

I	View Name	Description	SQL
7	v_ts	Area Query	select *,20 as f_e from ts_b {where}
5	v_partinout	The Query for Material In/Outbound	select a.f_id,a.f_name,syjc.f_num as
4	v_prodinout	The Query for Product In/Outbound	select a.f_id,a.f_name,syjc.f_num as
2	v_ck	Warehouse	select * from t_ck
1	v_td	Multi-person	GetTDData(select f_id,f_type,f_num from
8	v_partxqtoplan	Material Requirements Balance	select f_wlid,f_name,f_xqdate,f_ddnu from
7	v_prodxs_perdate	Sale Statistics (PD)	select b.f_name,b.f_type,b.f_spec,a.f_name from
6	v_prodxs_permonth	Sale Statistics (PM)	select b.f_name,b.f_type,b.f_spec,a.f_name from
5	v_prodxs_peryear	Sale Statistics (PY)	select b.f_name,b.f_type,b.f_spec,a.f_name from
4	v_prodxs_perkh	Sale Statistics (Client)	select b.f_name,b.f_type,b.f_spec,a.f_name from
3	v_partdd_wgh	Purchase Order for Non-Supply	select a.*,b.f_name as f_wlname,c.f_wlname
2	v_prodsc_perdate	Output Statistics (PD)	select b.f_name,b.f_type,b.f_spec,a.f_spec

View Structure | SpecMenu

Field Name	Field Title	Field Size	Field Widt	Field Orde	Show	Date Filter	List Filter	Total
f_date	Date	50	120	0	False	True	False	False
f_id	ID	50	126	1	True	False	False	False
f_name	Product Name	50	95	2	True	False	False	False
f_syjcnum	Unused Balance in La	0	78	3	True	False	False	True
f_innum	Inbound Number	0	69	4	True	False	False	True
f_outnum	Outbound Number	0	62	5	True	False	False	True
f_xyjcnum	Unused Balance in Tl	0	72	6	True	False	False	True
f_ksnum	Loss Number	0	93	7	False	False	False	False

stamping parts, sheet metal formula management, plastic parts management, customer-products relationship, materials-supplier relationship.

Warehouse Management

Storeroom management, inventory management, outside storage management, product input (plan), product output (other), product output (sales), product output (other), materials input (plan), materials output (other), materials output (planned), materials output (other), storage balance (inventory), statements of material, statements of products.

Warehouse Management is responsible for the warehouse data management of the company, including the interior warehouses, workshops and external storage warehouse.

Sales Management

Sales order management, shipping orders, sales returns management, sales invoices management, sales receivable management and accounts receivable statistics, sales statistics (per day /per month /per year /per customer).

Sales management is responsible for the management of sales data, which can inquire various sales data conveniently.

Production Management

Current plan, semi-finished plan, parts and components purchase plan, production plan, current demand for materials, demand for sheet metal, plastic demand, product demand balance, balance of demand and output statistics (per day / per month /per year).

Production management is responsible for the establishment and maintenance of various production planning; it is very convenient to establish a plan with the tools provided by the system.

Quality Management

Quality inspection of outsourcing material, association of quality and technical supervision, quality inspection of failed batch statistics (material/suppliers). Quality management is responsible for maintaining and analyzing the quality testing data of outsourcing material.

Procurement Management

Purchase order management, material return, purchase payment, payment statistics. Procurement management is responsible for the maintenance of purchase orders and payment.

Workshop Management

Workshop condition statistics, production journal record, daily statistical analysis, production overview, statistical analysis of staff production, staff incentives statistics, overtime registration, leave registration, overtime statistical analysis, attendance statistical analysis, process management, maintenance of failure reasons, setting leaving reasons.

The material accounting data flow chart of system is shown in Figure 18.

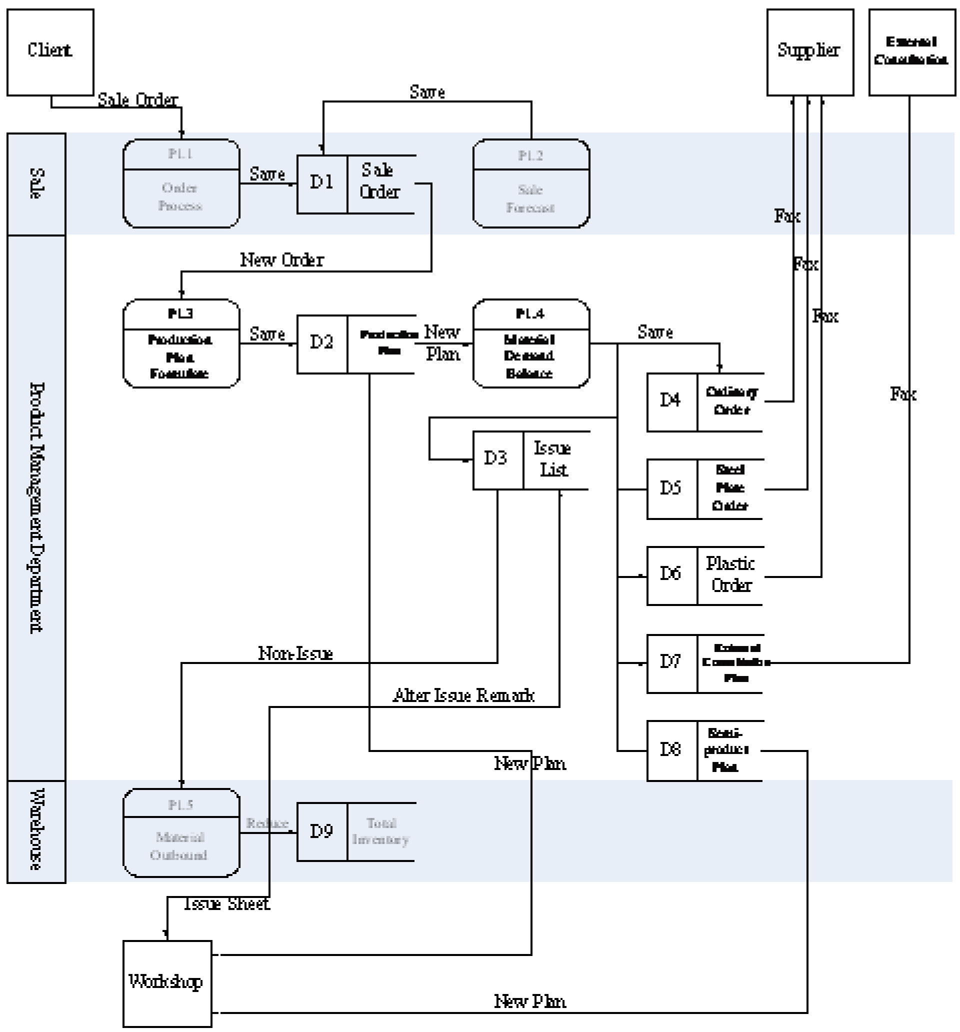
Since the ERP system has been used, the ability for adapting to the market has been greatly enhanced. The logistics, information flow and capital flow can be effectively managed and information is shared, which enhances the management level of the company. Specific benefits are as follows:

1. It has improved inventory management, reduced inventory by around 15 percent, increased the accuracy rate of in-store materials and finished product, improved input and output statistical efficiency in the store, reduced inventory checks error rate, and reduced error rate for checking parts storage, which has greatly improved the situation of cash flow and production supply.
2. It has improved labor productivity by approximately 18%, accomplished timely productiony, arranged the production progress reasonably, and increased the rate of the delivery schedule.
3. It has strengthened workshop management control and the control workshop production progress.
4. It has reduced the cost of production by about 12 percent; management staffs are able to know and analyze the production situation in good time, and he monitoring and management of failed production is strengthened.
5. The system has enhanced the efficiency of providing information to management and has strengthened the ability to adapt to the market and rapid decision-making.

Example II: ERP Systems for Optoelectronics Industries Based on ISO9000

Main function modules of the system include:

Figure 18. The material accounting data flow chart of system



Basic Data Management

Basic data management of staff, materials and products basic data management, materials inventory management, management of setting authority, equipment and fixture database management, data dictionary management and so on.

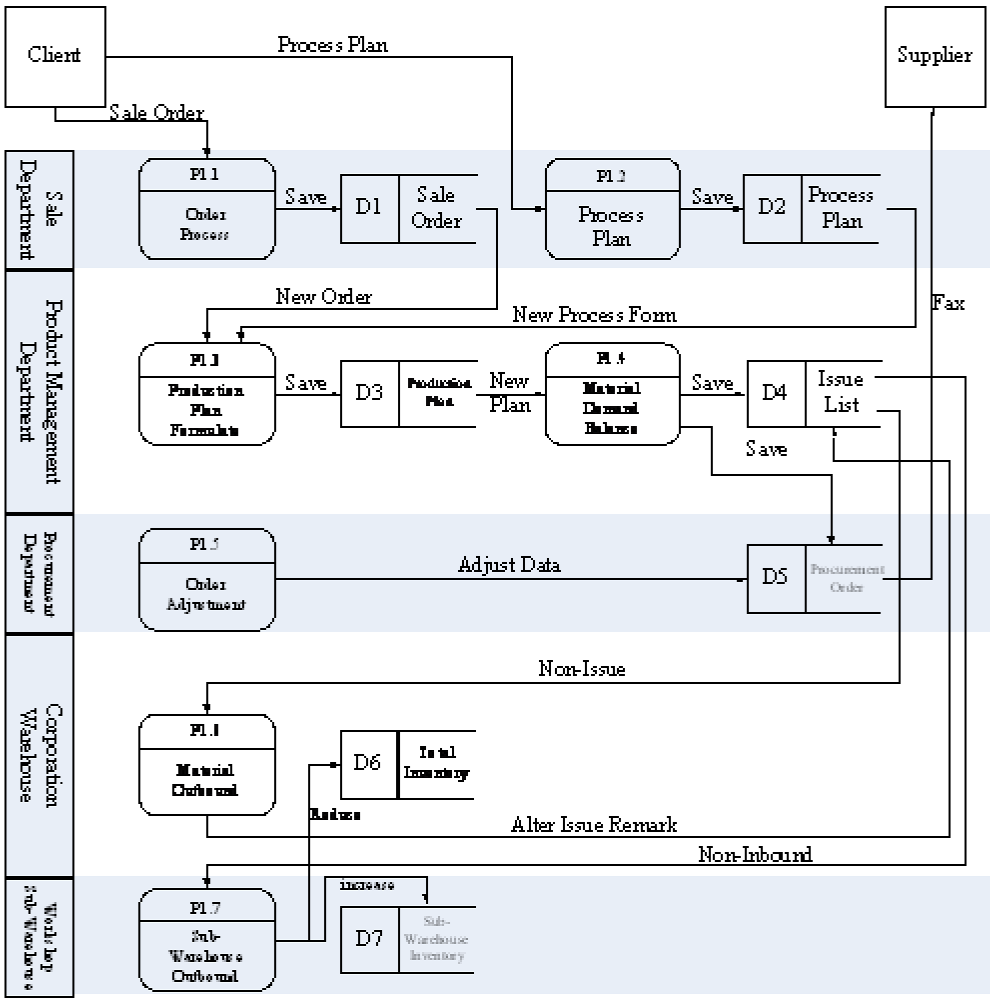
Marketing and Sales Management

Customer resource management, customer demand/feedback information management, sales order management and after-sales service support management, competitor analysis, and cost and contract management.

Quality Management

Quality planning and document management, equipment and testing measuring tools management, quality improvement tool sets, quality inspection statements, statistics, analysis and evaluation results of quality management.

Figure 19. The data flow chart of material counting part



Procurement Management and Supplier Management

Supplier resource management, supplier rating assessment, purchase order management, procurement quality of feedback management, procurement and contract management costs.

Production Management

Production management plan, on/offline data collection, product quality control of key processes, part-finished/finished product quality inspection management, diagnosis and control of process ability, statistical analysis for statements, quality improvement and decision support management, cost and quality control management.

Warehouse Management

The data flow chart of the material counting part is shown in Figure 19.

Material inventory management, part-finished/finished product inventory management, in/out management, inventory management and bal-

ances, occupancy management of warehouse cost, statistical analysis of inventory statements.

The system has operated in the company which was approved by some enterprises and has passed a project appraisal by the Foshan Municipal Science and Technology Organization. The system's operation has achieved the following benefits:

1. It has improved inventory management and increased the accuracy rate of in-store materials and finished products. The statistics efficiency of warehousing, delivery of cargo from storage and landed cargo has been improved. The inventory checks error rate and shortage of parts has been reduced. Inventory funds retention and production availability have been greatly improved.
2. It has improved labor productivity and rendered the production planning schedule more timely; it has arranged the progress of production reasonably and increased the rate of the delivery schedule.
3. It has strengthened the overall quality control of the product, realized statistics and analysis of the product quality inspection process, and ensured that the quality and quantity of production was linked to the operators' salary, which improved the level of staff responsibility and reduced the rate of failed products.
4. Integrating the system with financial data enhanced efficiency.

Conclusion

In this session, two cases has been used to demonstrate the novel Intelligent Management Information System for manufacture enterprises, The system not only has the feature of past enterprise management system, but also combines intelligent and reconfigurability in one. These cases has been described via three examples, which are Multi-agent IDSS based on blackboard, IR ERP systems for enterprises of Automobile Fitting and ERP Systems for Optoelectronics Industries Based on ISO9000. Reader will understand the basis and development of IMIS and its relative technology clearly by analyzing these two cases. The theory of these cases is based on the Chapter 2, 3, 4 and 5 of the book.

The new ERP concepts and technologies of the manufacturing process is important to the overwhelming majority of manufacture enterprises, because the specific feature, specially the reconfigurability that make the late-model ERP system better serve the requirement of enterprise management. The Intelligent Reconfigurable ERP (IR ERP) system will not only aim to enhance work efficiency by implementing ERP, but also to establish a comprehensive set of intelligent decision-making systems, which will transfer the focus of business management from transaction to high-level intelligent decision-making systems, such as management, control, planning and analysis of enterprise. The system is characterized by lower-level abstraction and a higher degree of abstraction. Not only does it achieve the separation of the interface and the logic of the implementation, but it also realizes the separation of the implementation logic and specific business rules, which realizes the completed parameterized design of the ERP system. Therefore, system development speed is greatly accelerated and flexibility is also enhanced. Helping the enterprise to realize the management innovation by information technology, improve the management level and enhance the competitive capacity.

SUMMARY

Decision-making Support Systems (DMSS) are computer-based systems that support individual or organizational decision-making processes. Recent advances in information technology and artificial intelligence continue to enhance these systems and have given rise to IDMSS. This chapter has provided integrated coverage of the

technical aspects of Intelligent Decision-Making Support Systems together with discussion of their application and evaluation in ERP. Readers will learn about the foundations, architectures, methods and strategies for successfully designing, developing, implementing, and evaluating intelligent Decision-making Support Systems. Intelligent Decision-making Support Systems: Foundations, Applications and Challenges will be of value to researchers in AI and management studies interested in the latest thinking in decision-making, as well practicing managers and consultants who are involved with putting advanced information technologies into practice in organizations.

Decision-making system research provides a powerful foundation and support for IDMSS. Decision-making system research is divided into eight fields: prototype design, method structure, conception design, experimental research, field testing, research summary, case study, conclusion and thinking. These fields provide the theory and practice basis for DMSS development. Research in these fields is still important to IDMSS study, however, the intelligence characteristic of IDMSS is different from DMSS in that it includes, composition and integration of intelligent decision-making, comprehension of the decision process, the space time and multi-dimension decision process, distributed parallel solution, and human-computer interaction based on knowledge. All five fields will be future directions of IDMSS.

REFERENCES

Bernroider, E., & Koch, S. (2001). ERP Selection Process in Midsize and Large Organizations. *Business Process Management Journal*, *7*(3), 251–257. doi:10.1108/14637150110392746

Corkill, D. D. (1991). Blackboard Systems. *AI Expert*, *6*(9), 40–47.

Corkill, D. D. (2003, October). Blackboard and Multi-Agent System & The Future Collaborating Software. In *Proceedings of the International Lisp Conference*, New York.

Elena, Y. G. (2002). An Approach to Designing Intelligent Decision-Making Support Systems on The Basis of The Logic-Semiotic Apparatus. In *Proceedings of the 2002 IEEE International Conference on Artificial Intelligence Systems (ICAIS'02),* September.

Er, M. C. (1988). Desionon Suppon Sydems: A Summary, Problems and Future Trends. *Decision Support Systems,* Apr.

Fan, J. C., Liang, Y. Q., & Zeng, Q. T. (2006). Method for Designing Organization Decision Support System Framework. *Journal of Systems Engineering and Electronics*, *17*(4), 764–768. doi:10.1016/S1004-4132(07)60013-9

Gao, M. M., Zhou, M. C., & Reggie, J. C. (2003). *Optimal and Intelligent Decision Making in Sustainable Development of Electronic Products*. Newark, NJ: New Jersey Institute of Technology.

Guo, S. S., Yang, M. Z., Zang T. G. (2003). Research and Development of A Network-Based Intelligent Maintenance Information System. *International Journal of Plant Engineering and Management,* (02).

Hu, S., & Wu, H. (2004). A Decision-Making Model for Intelligent Agent with Internal State. In *Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04), IEEE Computer Society*, (Vol. 02), September.

Huang, H. P., & Liang, C. C. (2002). Strategy-based Decision Making of A Soccer Robot System Using A Real-Time Self-Organizing Fuzzy Decision Tree. *Fuzzy Sets and Systems*, *127*(April), 49–64. doi:10.1016/S0165-0114(01)00152-X

Huang, J. G., Zang. T. G., Zhang, S. M. & Yang, M. Z. (2003). Study on The Decision-Making Support System of Intellingent Maintenance. *Mechanical Engineer,* (12).

Jatinder, N. D., Forgionne, G. A., & Mora, M. (2006). *Intelligent Decision-making Support Systems: Foundations, Applications and Challenges (Decision Engineering)* (1st ed., pp. 50–55). Berlin: Springer Press.

Jiang, Y. C. (2006). Research on Intelligent Decision-making Support System. *Boiler Manufacturing,* (01).

Jolly, K. G., Ravindran, K. P., Vijayakumar, R., & Kumar, R. S. (2007). Intelligent Decision Making in Multi-Agent Robot Soccer System Through Compounded Artificial Neural Networks. *Robotics and Autonomous Systems Archive*, *55*(7), 589–596. doi:10.1016/j.robot.2006.12.011

Kim, C. J. (1989, January). *An Intelligent Decision-Making System for Detecting High-impedance Faults*. Texas A&M University.

Liu, R. Zheng, W. & Wu, Y. (2006). Intelligent Decision Making for Agreement-based Grid Resource Management. In *Proceedings of The First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06),* (Vol. 01).

Manheim, M. L. (1998). An Architecture for Active DSS. In *21st International Conference on System Science*, Hawaii, (pp. 356-365).

Phillips-Wren, G., Ichalkaranje, N. & Jain, L. C. (2008, April). *Intelligent Decision Making: An AI-Based Approach (Studies in Computational Intelligence)* (pp. 44-50). Berlin: Springer Press.

Tu, X. (1992). Intelligent Management System. In *The 7th International Conference on Modeling and Control of National Economics*.

Xu, D. L., & Yang, J. B. (2003). Intelligent Decision System for Self-assessment. *Journal of Multi-criteria Decision Analysis*, *12*, 43–60. doi:10.1002/mcda.343

Yang, A. (1990, January). *An Intelligent Decision Support System for Decision-Making Based on Vague, Imprecise and Incomplete Information*. Purdue University, West Lafayette, IN.

Yang, B., Ma, Y. & Yu, L. (2004). Intelligent Decision Support System for Bank Loans Risk Classification. *Journal of Donghua University,* (02).

Yang, M., Huang, J., & Zang, T. (2004). Research on An Intelligent Maintenance Decision-Making Support System. *International Journal of Plant Engineering and Management*, (02).

Zhang, R., Sun, J. (2006). Study on Intelligent Spatial Decision Support System of Agriculture Study on Intelligent Spatial Decision Support System of Agriculture. *Journal of Harbin Engineering University*, (z1).

Zhu W., Yang S., & Ren, M. (2002). Intelligent decision support system: State of art and challenges. *Journal of Systems Engineering,* (05).

Chapter 14
Trends and Prospect of Manufacturing Intelligence

Manufacturing Intelligence (MI) focuses on how to adopt artificial intelligence and computing intelligence methods to solve problems in digital manufacturing, such as intelligent digital scheduling, intelligent digital designing, intelligent digital machining, intelligent digital controlling, intelligent digital process planning, intelligent digital diagnosis and maintenance. In order to meet the needs of a changing market, manufacturing systems have been in a constant process of development and the pursuit of perfection. In this chapter, the various characteristics of the manufacturing industry in the 21st century are analyzed from the perspective of society, market, product, enterprise and manufacturing technology. Then a brief review of MI in previous chapters is presented. The trends and prospects of MI, adapted to the development of manufacturing, are then discussed, followed lastly by a summary of this chapter.

DOI: 10.4018/978-1-60566-864-2.ch014

INTRODUCTION

The manufacturing industry is the mainstay of any modern economy and a reflection of national strength, in which the gross production value generally accounts for 20%-55% of a country's gross domestic product (GDP). In most countries, manufacturing technology usually plays a 60% role in the structure of enterprise productivity (Li, 1998). The competition among national world economies lies mainly in the competition of advanced manufacturing technology, and this competitive is finally embodied in the market share of products (Kahn, 2001). With the rapid development of economy and technology and the equally rapid changing of customers' needs and market environment, this competition becomes more intense, and governments attach a great deal of importance to the research of manufacturing intelligence technology. With the advancement of science and technology, and the development of society and economy, 'manufacture'

has evolved from a simple process finished by an individual worker and an independent machine to a complicated project which has to be finished by a manufacturing system composed of many manufacturing elements. From the perspective of system science and engineering, the manufacturing system which undertakes and completes the modern manufacturing task has developed from a simple system to a complicated one.

From the perspective of the history of manufacturing not only is the scale of the manufacturing body (for instance, the factory) continually enlarging, but its structure is also becoming more and more complicated. In the times of handcrafted industry, the designing, machining, assembling and detecting were finished individually; the scale of the manufacturing system was small, the structure was simple, and it was easy to operate. During the industrial revolution in 18th century, manufacturing systems were still quite small, although manufacturing technology and related equipments had developed considerably. Throughout the mass production times of the late 19th century, operations were specialized through labor division and, with the support of mechanism and electronics, manufacturing systems on a large scale were developed by using an assembly line. Nevertheless, the structure was not complicated, and it was comparatively easy to cope with problems such as analysis design, information process and management control. In the second half of the 20th century, however, particularly over the last 30 years, the diversity of market needs and competition have forced product manufacturing to develop in diverse directions, with different levels of production and short product cycles. As a result, advanced manufacturing modes such as soft manufacturing, computer integrated manufacturing and sensitive manufacturing have been created, and the manufacturing system is not only very large in scale, but also has a complicated system structure, material flow and information flow. Furthermore, the analysis and integration of the system faces many difficulties, as it is much harder to deal with optimization management control problems, which has become a challenging problem around the world (Dimopoulos, 2000).

With the development of computer science in the 1960s, many new control technologies and methods were applied to engineering projects and products, which greatly enhanced the development of industry technology (Paul, 2002). In the late 20th century, the rapid development of burgeoning science technologies and their application in manufacturing, represented by microelectronics, information technology, new material technology, artificial intelligence technology and system science, greatly expanded the depth and range of manufacturing activities. These changes also impacted on design methods, product structure, production mode, production process and production organization of modern manufacturing and lead to profound changes in expressing, storing, transmitting and processing manufacturing information, and saw the emergence of a number of novel manufacturing technologies. Digitalization and networking have acted as indispensable driving forces in the life cycle of the manufacturing product. In this new manufacturing technology environment, digital manufacturing has emerged to become a strong driving force in promoting development of the manufacturing industry in the 21st century. As the information age marches towards the intelligent age, intelligent digital manufacturing systems, integrated automation (Lastra, 2006), networking and intelligence, embodying the most advanced technologies, have promising prospects and will form the intelligent productivity of the new generation.

DRIVING FORCES AND CHALLENGES OF THE MANUFACTURING INDUSTRY

Changes in the Manufacturing Environment

The intelligent manufacturing system is quite different from the traditional manufacturing system and responds to various customer needs. It is open system, and its advanced characteristics include changes to materials, energy and information; with changes in demand, it has the ability to reconfigure its system in order to adapt to the changes in the environment. It is therefore necessary to thoroughly research the manufacturing industry environment when studying the intelligence system. This environment includes both the social and market environments.

Three Characteristics of Social Environment Change

1. **Industrial economy creates issues within the context of social development.** While development of the industrial economy creates a modern civilization and vibrant success, it is also brings many problems of socially sustainable development, that is, the three worldwide problems of resources, environment and population. It manifests in the following three aspects: limited resources, worsening environment and rapid growth of population and social aging process. The overuse of resources has induced a crisis-survival situation for humans, and the shortage of resources becomes increasingly obvious (Ward, 2000); for example, worldwide energy consumption in 1996 was six times that of 1990. The ecological environment continues to deteriorate, and every year, the manufacturing industry produces tonnes of waste. In 1996, 24 million cars were discarded; in 2000, 20 million computers were thrown away.
2. **The knowledge-based economy has become the new trend of world economy development.** Invention of the engine, electricity and the birth of the automobile brought human beings into a time of industrial economy, and the invention of the computer and the application of information technology delivered the whole world into a knowledge-based economy (Wayne, 2002). In 1990, the United Nations research organization puts forward the concept of a 'knowledge-based economy', and defined the characteristics of this kind of economy. In 1996, the Organization for Economic Cooperation and Development (OECD) puts forward 'Knowledge Based Economy', in which human capital and technology are at the core of economic development. A knowledge-based economy is one that is based on the production, delivery and consumption of knowledge. Knowledge not only includes the social, arts, nature and technology sciences but also includes the production and accumulation and application of knowledge. Knowledge is an unlimited resource, which can have far-reaching effects on economic development, and its value as a resource is greatly enhanced when it is shared and widely transmitted (Mabert, 2003). Today, the value and price of a new product mainly depends largely on its knowledge and technology content, and the exclusiveness of its technology (Koh, 2006). This bears little relationship to the direct cost, and the development of future economies relies largely on the impetus of knowledge and technology. A knowledge-based economy is an economic mode that is superior to the industrial economy. The three characteristics of a knowledge-based economy are that it is an economy guided by an information stream within a mature information network; an economy which takes scientific knowledge and talent education

as its two pillars; and an economy of open cooperation and sustainable development. In some developed countries, the knowledge-based economy is beginning to replace the industrial economy. According to survey, during the ten years from 1987 to 1996, the share of high-tech products of OECD membership countries in manufacturing and export doubled to 20%-25%. To date, more than 50% of the economic development of the principal OECD countries is knowledge-based. As some experts estimated, the contribution of scientific progress to economic development has risen from 5%-20% in the early 19th century to 70%-80% during the 1970s to 1990s.

3. **Global economic trade has intensified.** In the last twenty years of the 20th century, the World Trade Organization(WTO) was founded, the European union instigated its euro plan, and the Asian Pacific Economic Cooperation (APEC) and some economic unions in Northern America and South Asia were developing rapidly, all of which greatly increased the speed of globalization. With better compliance with WTO rules and the building of global transportation and communication, international economic cooperation and contact becomes closer, global industries enter an important period of structure-adjustment, and the world becomes a bigger and more unified market. The globalization of the manufacturing industry, manufacturing products and manufacturing technology results in redistribution and reconstruction of the manufacturing industry in the global market, and makes economic competition worldwide increasingly serious. Industrially developed countries enhance their manufacturing competitiveness through the reconstruction of enterprises and the use of high technology, while the manufacturing industry in some developing countries continues to develop rapidly. The localization and globalization of production, supply and sales exist side by side and frequently clash with each other.

Five Characteristics of the Changes of Market Environment

Market changes are closely related to social requirement, productivity, scientific progress and economic development. Today's market environment includes the following five characteristics,

1. **Buying market.** As a result of scientific and technological development, the 'seller's market' has gone forever. Buyers have more choices for their products, and may there have some strict requirements for the products that sellers provide. Besides traditional requirements such as low price and high quality, buyers now pay more attention to shipping dates and after-sale service. Enterprises must be customers-oriented (Kwon, 2007) and the Time Quality Cost Service Environment (TQCES) has become a decision property of the new generation of manufacturing system.
2. **Uncertain market.** Scientific development, variety of demand and changeability result in a fast turnover of product generation, resulting in a great demand for non-mass production, decentralization, and customization. As the market becomes larger and larger, competition becomes more fierce, uncertainty factors increase and the market changes faster.
3. **Global market.** As the global market becomes more open, the market becomes global. Countries seeking trade protectionism will be undoubtedly excluded from the world economy family. As a result rising third-world economies such as Brazil, China and India, global market structure has

changed significantly, and the globalization of industry division has been a development trend.

4. **New market.** New social requirements brought about by social ageing, agricultural modernization, environmental protection, the modernization of living conditions, renewable resources, better city traffic and waste solutions require new products and new industries. This not only involves new products which are combination of traditional and high-tech products, and but also involves new markets that previously did not exist, such as the development of the software industry as a knowledge-based industry and the development of the 'green' industry (King, 2001).
5. **Virtual market.** The further development of the information manufacturing system is the network manufacturing system, in which online product advertisement, product display and exchange, client relationship and agents all belong to virtual market.

Changes within the Manufacturing Industry

Today, manufacturing industry, as the pillar industry of one industrial country, is not traditional mechanism, instead, it is a large manufacture industry which contains lots of fields. Every manufacturing enterprise has to face the whole process of product life cycle, and solving so many problems needs high-tech of multi-fields, which means the complexity of manufacturing technology, the variety of manufacturing mode and complication of manufacturing system, meanwhile, it is one of challenges when manufacturing industry takes reform, restructure .

Six Characteristics of Product Changes

Products are something that are provided to the market to attract people's attention, generate consumption and to satisfy people's desires and needs. Since consumption is the original driving force of production, so market changes affect and determine product changes.

1. **Product connotation increasing.** Product is a set of solution projects that meet customers' need, and it is also a system. It contains entity, software and service. Among them, the proportion of latter two is increasing, and the service function of manufacturing industry as well. Service includes the whole process from product development to sales and after-sales service. To satisfy customer need is the result of a 'buyer's market'. Initially, the requirement was that products should be of good quality and low price; then short shipping dates, good service and a friendly environment were added; now another element is added, that is full of technology and knowledge.
2. **Shorter product life cycle.** This is embodied in two aspects: a shorter design and manufacturing period before shipping to customers, and a shorter use period for customers (Cheng, 2001). The reasons are as follows: the advancement of science and technology; the pressure of competitive environments forces producers to respond quickly to changes in the market; the competition strategy for creative products; a shorter product market life cycle can speed up new product development, as shown in Table 1. IC(Integrated Circuit), for example, its integration level double times and the price halves, every 18 months. Optical technology develops a new generation every 21 months. The core competitiveness of enterprises in the marketplace hinges on developing new products to satisfy customer need at the fastest speed and putting them to market within the shortest possible time.
3. **The variety of products.** The product types is becoming more and more, and it can be

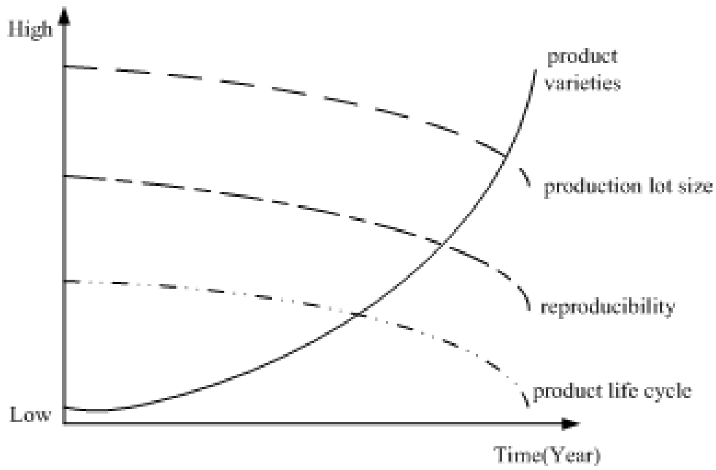
Table 1. Instance for shorten development period in manufacturing industry

Year	Product Market Life Cycle(Year)			Development Period (Year)		
	Automobile	Electronic Product	Machine Tool	Automobile	Electronic Product	Machine Tool
1990	7.8	13	14.5	2.0	2.5	1.9
2000	2.5	8.7	5.5	1.3	1.3	1.3
Reduce Ratio(%)	78	33	63	35	48	31

called" the product is everywhere". The demand for unique products requires that products should be diverse in style, function characteristics, art appreciation and service culture, as shown in Figure 1. A single one-size massive consumer market is gradually divided into a set of smaller markets that satisfy different needs, and these smaller markets further strengthen the variety of products.

4. **Enhancement of product intelligence.** Products, in nature, not only replace, strengthen and extend man's labor work, but also contain some 'intelligence' because of digitalization, that is, the information recognition function, information process function, information store and display function as well as the combination of all functions. This characteristic is the result of mature intelligence technology and its wide use. Thanks to the IT and service industries, products combined with machines, electronics, light, measuring, control and information are increasing in the market. Examples can be found in multi-media education and entertainment equipment, networks and equipment, electronic finance and business equipment, mobile communication equipment, GPS, minor mechanical and electronic devices, life project devices and limiting environment operation.
5. **Better compatibility of products.** In order to meet the need of global competition, cooperation, interconnection, and product upgrade, keeping technological advantages and characteristics, taking standard interface and compatibility technology can realize global interconnection and technological compatibility (Shah, 2003).
6. **The ecological and environmental property of products.** Consideration of environmental issues in the production and consumption of products is becoming increasingly important Green products, during their life cycle, save energy and resources, do not pollute the environment and are convenient to recycle or reuse. Population aging and the consequent need for increased health-care, medical detection, treatment, monitoring and gym equipment, and the enhancement of environmental protection awareness, and renewable resources will undoubtedly become new profit points for the manufacturing industry. The development of new material is an attractive concept; practically, new material not only enhances product function, but also creates new products and new processes, and the manufacturing industry will increasingly be required to consider material consumption, composite material application, material application, environment-friendly material development and material renew.

Figure 1. Product changes with time



Six Characteristics of Enterprise Changes

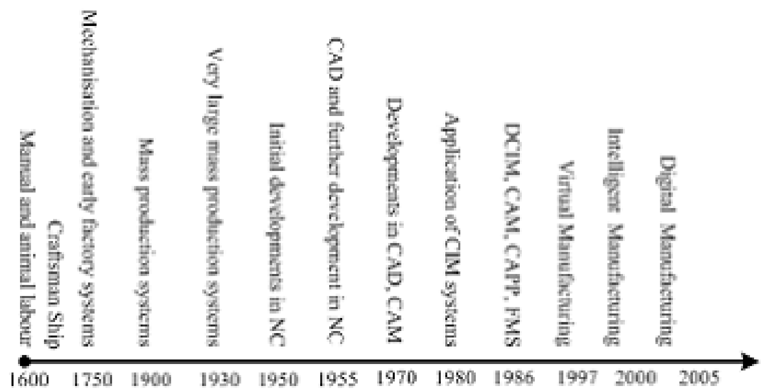
Today's enterprises are experiencing all-encompassing changes, which are shown as follows.

1. Changes in enterprise strategy. The greatest change here is from a quality-first strategy to a strategy of fast response to market need. To meet the customer's changing product requirements, fast response is necessary, so shipping date has become a key element of competition; the rapid restructuring of production requires the manufacturing mode to be highly 'soft' and sufficiently sensitive, in response to this need for 'customization'. Information technology and management science provide important guarantee for this change. For example, division of operations divides each enterprise into several self-managed units which can be responsible for themselves. In this way, an 'inside market' is formed which is dynamic and has strong adaptability to environmental changes, thus creating a highly-effective complex.
2. The change of relationship from competitors to cooperators. Globalization will result in the inter-development of localization and globalization. A philosophy of 'you die' or 'I lose' will only result in a negative outcome for both parties. Instead, 'win-win' should be the result of cooperation in the market. Cooperation is an important trend in modern enterprise activities, and allows knowledge to be shared on a larger scale, and so improve efficiency. Networking provides a good, convenient condition for cooperation.
3. Change of making profit from extensiveness to intensiveness. Profit increases in the past have largely been achieved by investing more money and using more resources and manpower, but now, this is more likely to be achieved by developing knowledge resources. Knowledge resources gradually become the major strategy resource for economic development, and knowledge management (KM) is firmly on the agenda. In a time when science and technology is developing so fast, enterprises will undoubtedly meet many setbacks and possibly fail if they only emphasize equipment and technology and ignore human resources and human spirit. KM uses group power to enhance adaptability and creativity, to provide new ways

for enterprises to share visible and invisible knowledge. Visible knowledge is that which is easy for a computer to process and store; invisible knowledge is harder to master, since it comprises the personal experience and skills contained within every employee's mind.

4. Change of operation mode from production manufacturing enterprises to development operation enterprises. Since the 1990s, new product development has been the lifeblood of enterprises. With faster speeds of information transmission and faster development of new products, enterprises attach more importance to building up an ability to exploit the global market, including a sensitivity to requirements, market modeling and requirement report, timely market information base building, promotion strategy, pre-sale and after-sales service satisfaction, online business exploits, building of virtual enterprises and virtual supply chain, only the combination of natural resources and advanced technology create more profit.
5. Change of working mode from traditional sequence working mode to parallel working mode. Parallel project extend to parallel manufacturing. Parallel manufacturing can obviously shorten the time from product concept to final product. In cooperative firms, the core competitiveness and knowledge of different sections of each outside firm can be dynamically restructured, to reduce investment risk largely through accurate computing, optimization and tracking product cost (Duvivier, 2003; Yoshimura, 2007). Parallel manufacturing brings revolutionary changes to research, development and manufacturing modes of different layers; parallel manufacturing not only needs important new technology in communication and data-processing, but also requires a new social and cultural atmosphere for manufacturing enterprises. This is extremely important for a global, multi-disciplinary, multi-culture and fast-response organization. The multi-department, multi-layer, organization structure of the past is changing to small-group, self-managed, dynamic and networked organization structure.
6. Change of manufacturing technology from researching traditional process mechanisms and reforming traditional process technology to exploit a manufacturing method that processes by limiting conditions and developing advanced techniques and new devices. Technique and device are the core of manufacturing technology, as well as an embodiment of an enterprise's competitiveness. In order to exploit new products and new manufacturing technology, more sophisticated, smaller, faster, better soft quality, more stable and less polluted techniques and devices should be invented. The main changes of manufacturing are typified in the following three aspects.
   a. Development and application of advanced techniques, including the sophisticated update of traditional technique, unit technique and its formation technology, the special material process, minor process, super sophisticated process, super high-speed process, less or no cutting fluid process, less or no clamp process, intelligent tool and knives, and high energy process.
   b. Research for new devices. This consists of the following three aspects: structure creation (such as numerical control, virtual axis machine tool), function creation (such as robot lathe, standard soft manufacturing system, and restructured manufacturing system) and repair creation (including equipment error report and equipment non-error operation technology).
   c. Closely combined with information technology. This mainly includes:

Figure 2. Evolution in manufacturing technology



popularizing CAX and modeling and emulation technology; using virtual technology to speed up enterprises' activities, improve product quality, save costs and respond quickly to change; using intelligent systems to serve the whole system and make it easy to communicate and work.

Development Trends of Manufacturing

Manufacturing is the general name given to a series of activities and work involved in the manufacturing industry, such as product design, material selection, production planning, production processes, quality assurance, management, market sales and service (Madhusudan, 2005). In the last 40 years, information technology has developed at high speed, and the fusion of information technology and manufacturing technology has resulted in the digitalization of manufacturing (Chisholm, 1990). Manufacturing technology develops rapidly, and the manufacturing system has changed from an energy driven mode to an information driven mode, as shown in the evolution of manufacturing, Figure 2. Digitalization has been an indispensable driving force in the life cycle of product. Digital manufacturing is a novel manufacturing mode, adapted to complicated production structures, consumer demand and giant manufacturing networks, and has naturally become the major characteristic of the future of manufacturing industry development (Mehrabi, 2000). Digital manufacturing utilizes digitalized quantitation, expression, storage, processing and controlling methods to support the global optimization operation in the product life cycle, characterized as digital modeling simulation and optimization, taken the knowledge fusion as the basis; it is supported by virtual reality, computer networks, rapid prototyping, database technology; it analyzes, plans and reorganizes the product information, processing information and resource information according to the client's requirement, and implements product design, function simulation and prototype manufacturing; it is a whole manufacturing process that rapidly produces a product to satisfy the client (Zhou, 2006). In fact, digital manufacturing is a whole manufacturing process that produces a product in the digital space constructed by digital description. The kernel of the digital manufacturing system lies in three aspects: digitalization, virtualization and networking.

1. **Digitalization.** Manufacturing digitalization is not only the result of crossover, fusion, development and application of manufacturing technology, computer technology, network technology and management science, but

also the inevitable trends of the manufacturing system. It begins with the digitalization of product design, and permeates into various activities, equipments and entities in the product life cycle (Sun, 2001). It contains three main parts: digital design, digital control and digital management. Compared with 'analog', 'digitalization' in information processing, has the advantage in precision, security and high capacity. The control parameters of the manufacturing equipment are all digital signals. Product information, process information and resource information for enterprises are all transmitted in digital form over networks, assuring rapid reconstruction and fast response to the market and meeting the client's requirements. In the global manufacturing industry, the client publishes information over networks; diversified enterprises then form dynamic alliances through the network, to design and produce the corresponding product collaboratively and rapidly, according to the client's demands. In the digital manufacturing environment, enterprises, warehouses, workshops, equipment, staff, distributors, and even relative markets are considered to be 'nodes' in the wide ranging digital manufacturing network. In the process of research, design, production, sales and service, each node in the manufacturing network shares its resources, centering on the digital information given by the product, and participates as the most active factor driving the manufacturing industry.

2. **Virtualization.** Taking research on the manufacturing process as a category of the micro process digitalization, the layout design and actual optimization operation of the producing system is classified into the field of macro manufacturing process digitalization. The key to virtualization is computer simulation. Virtual manufacturing can be simply comprehended as 'manufacturing inside the computer', which generates a soft prototype substitute for the conventional hard prototype through the simulation software system. The system simulates, builds a model and analyzes the performance of the product, process and enterprises to assure the reasonableness of product design and process and the success of the product manufacturing and production cycle, and finds the inevitable defects and errors in the design and production process (Rosenzweig, 2003). The virtual technology includes virtual environmental technology, virtual design technology and virtual manufacturing technology (Mok, 2001). Digitalized description is the basis of virtual manufacturing. The unified modeling of procedures like designing, processing and assembling constitutes the virtual environment, virtual process, virtual product and virtual enterprise. It implements the virtual process in the following three layers. (i) Product design. The virtual design and assembly are implemented in the design process; the client participates in the digitalized product modeling by 'deep interactive' and immersive parallel development in a virtual environment, fulfilling the aim that lets the client experience the performance of the product in advance (Yuan, 2003). (ii) Production inside the factory. The machinability, processing method and process rationality of the product is tested in the production layer, which assures an effective and timely evaluation of the manufacturing system performance. (iii) Collaborative outside factory. Modeling and simulation of the manufacturing plan, management, shop scheduling, supply chain and logistics is performed, and a virtual research and development center and virtual enterprise are built (Wang, 2004). The virtual research and development center combines an excellent research and development team in diversified areas through

the network and video system. In order to respond rapidly to the demands of market, a collaborative economy entity, virtual enterprise, is constituted temporarily by different enterprises involved with the product, which transcends spatial constraints, and connects over network and to unified commands. Its features are virtualized functions, virtualized organizations and virtualized regions (Wang, 2006). Virtual manufacturing is the inevitable result of various information technology developments and the demands of modern enterprises.

3. **Networking**
   a. **Networking is a necessary road for the development of enterprises.** Under the great pressure of fierce competition, each enterprise has to implement profound reforms of the manufacturing organization, and concentrate on its own most competitive core business. Fortunately, this reform is made possible by computer technology and network technology. In network manufacturing (NM), new network space combines closely with the traditional space to generate diversified new thinking, new opinions, new methods and new systems. A new manufacturing mode, i.e., virtual manufacturing organization, 'virtual enterprise' and dynamic 'enterprise alliance' emerge. Each enterprise in a virtual enterprise concentrates on its own core business, complements advantages, and implements resource optimization, dynamic combining and sharing.
   b. **Content of network manufacturing.** This includes manufacturing environment networking, manufacturing enterprise networking, networking among enterprises, and the networking market.
   c. **Distributed collaborative design adoption of grid technology.** Grid technology is the main method of implementing resource sharing and integration between enterprise and society, and of providing an integrated supporting environment to the enterprise group's collaborative operation and management. Currently, the main application of the grid is to integrate the dispersed computing ability. In the future, the grid will greatly satisfy the demand of network manufacturing. Enterprise, even individual, will obtain manufacturing service from network, as convenient as acquisition of water, electricity, gas and Internet information, in the future.

REVIEWS ON FOREMENTIONED MI TECHNOLOGIES

Over the past several years, the intelligent technologies have been widely used in manufacturing aspects such as machine learning, planning and robotics, modeling human performance, expert systems, automated reasoning, etc. Many of these systems, such as advanced manufacturing systems (AMS), computer integrated manufacturing (CIM), flexible manufacturing systems (FMS), manufacturing resource planning (MRPII), CAD/CAM, numerical control machines are being developed for production and operation management and present a cross fertilization of ideas from MI. Several MI technologies have been depicted in the previous chapters. And a brief review of these technologies will be presented in this section.

Knowledge Based System

As one of the MI technology, knowledge-based system, especially expert system, has played an important role in manufacturing, and been recog-

nized as a promising paradigm for the next generation manufacturing systems. KBS is typically domain-specific and tends to operate by utilizing heuristic problem solving approaches. It provides systematic methods for the knowledge-based information processing in manufacturing. When a problem to be solved occurs in a manufacturing system, it is analyzed first and inferences are executed using inference rules and data stored in the knowledge base. The knowledge of experts is acquired and stored in the knowledge base in adequate form with the aid of the supporting system. However, knowledge is more than a mere collection of data, the term 'knowledge' carrying with it the concepts of 'structure' and 'understanding' (Halevi & Wang, 2007).

In chapter 2, an overview of the KBS is presented. Following the review, it will be seen that KBS is addressing a wide range of practical tasks, in various manufacturing fields, such as design, process planning, production control and diagnosis.

The KBS can be grouped into three broad classes depending on the nature of the problems that they solve (Pham & Pham1988). The first is 'derivation' (or 'classification') problems which include selection of machine elements, tools, equipment and processes, signal interpreting, condition monitoring, fault diagnosis, and machine control. In such problems, the problem conditions are stated as parts of the description of the solution. All the possible outcomes are known and the knowledge base is used to complete the solution. The problem-solving process could be viewed as the identification of a path leading from given conditions to desired outcomes. The second is 'formation' or 'synthesis' problems which cover all design activities (for example, component design, tool design, system design) where the solution space is completely open (infinite) and problem conditions are specified as properties that a solution must possess as a whole. Usually, there are several alternative outcomes, none of which is known a priori. And the last are problems such as process planning, production scheduling, and system configuring, which have features belonging to both the derivation and formation types. That is, the solution to these problems generally consists of combining elements drawn from a finite set. Although the outcome of the solution process thus implemented might not have been contained in advance in the knowledge base, all of its elements are.

The development of expert systems in manufacturing should be aimed at realizing more sophisticated manufacturing. The know-how which is the base of expert systems must be stored and expanded by the individual company, which is hard to get. However, once the know-how is stored in expert systems, it is invaluable to the management of MI.

Multi-Agent System

Agent theories and applications have appeared in many scientific and engineering disciplines. Agents address autonomy and complexity: they are adaptive to changes and disruptions, exhibit intelligence and are distributed in nature(Monostori, 2006). In this setting computation is a kind of social activity. Agents can help in self-recovery, and react to real-time perturbations.

Modern manufacturing system is a highly decentralized manufacturing system, consisting of a wide range of manufacturing resources, such as standardized or non-standardized, autonomous and semi-autonomous processing equipments, material-ferrying equipments, robots and so on. In this context, MAS are vital to realize important manufacturing properties as autonomy, responsiveness, redundancy, distributedness, reconfigurability and openness.

Agents could be designed to work with uncertain, incomplete information and knowledge. Hence, many tasks related to manufacturing, from engineering design to supply chain management, could be conducted by agents, small and large, simple and sophisticated, fine and coarse, grained

that were enabled and empowered to communicate and cooperate with each other. Beyond the usual advantages of having a modular system structure, additional merits are as follows: the agents need not be co-located, and they can form wrappers around and provide interface for existing legacy systems (e.g., analytical tools, simulators, CAD systems from various fields of engineering). Agents embodying knowledge and encapsulating tools of different engineering domains can communicate and work together if they have a set of shared concepts and terminology, a common language for expressing this knowledge, and a communication and control protocol for requesting information and services.

The application of MAS in manufacturing industries can manifest itself in the following aspects: production and information management; manufacturing process control; robot technology and parallel intelligent design and CAPP design. In recent works, CAPP has been extended towards the execution of plans: planning is aware of available resources and takes actual lot-sizes or due dates into consideration. On the other hand, process planning knowledge is used for short-term scheduling decisions at the shop floor. MAS concepts are particularly well suited to this integration.

The applications of MAS to above fields in manufacturing bring profound changes, like substantially enhanced the flexibility of companies, shorter lead times and associated higher production rate, reduction in waiting time of parts for service and lower variable costs, etc.

Computational Intelligence

The field of Computational Intelligence (CI) is mainly the result of an increasing merger of Fuzzy Systems (FS) or Fuzzy Logic (FL), Neural Network (NN) and Genetic Algorithms (GA). These technologies are providing increasing benefit to manufacturing industry(Mircea, 2005). However, each of them has own characteristics.

FS, which assimilate the concepts from fuzzy set theory and fuzzy logic, deals with the handling of imprecision in data, provides a framework for handling commonsense knowledge represented in a linguistic or an uncertain numerical form, finally makes a decision made by fuzzy concepts and human like approximate reasoning. There are two main characteristics of FSs that give rise to their popularity. Firstly, FS have knowledge bases that, because they are based on fuzzy IF-THEN rules, are easy to examine, understand, update and maintain. Another typical advantage of FS is their high degree of flexibility in dealing with incomplete and inconsistent data by aggregating the hypothesis of all the rules, especially for systems where a mathematical model is difficult to derive. The scope of successful FS applications is large and diverse. To name just a few, in industry with different intelligent microcontrollers; transportation with the control of rail, highway and air traffic and in financial banking with exchange prediction or stock market analysis. But FS also has real limitations– the heavy reliance on human experts in building the fuzzy IF-THEN rules, the need for manual specification of membership functions and fuzzy rules, as well as strong restrictions in automatic adaptive learning capabilities in an ever changing external environment.

NN deals with inherent parallelism in data flow. Their capacity to learn patterns in data that are noisy, incomplete and even contradictory, confer on NN two distinct advantages in processing information – the ability of processing incomplete information and generalize the conclusions. The application of NN is even larger in scope than FS. Just a few of the applications of NN in manufacturing industry are listed: signal analysis, noise elimination, data compression, quality control, production planning, diagnosis and training of system designers such as NC, CNC and DNC machines. NN effectiveness is due to the easy modeling and forecasting of non-linear systems and its ability to handle tasks involving incomplete data, unavailable expert advice or

where no rules can be formulated. From the machine learning point of view, NNs are capable of making non-linear mapping from a set of inputs to a set of outputs. They can be employed as universal functional approximators, which can offer accurate approximation of an unknown system on provision of data samples. The same as FS has its limitations, so does NN. NNs are unable to explain how they arrived at their conclusions and are unable to interact with conventional data bases. Also there is a lack of structured knowledge representation. Also NN have real scalability problems in that they experience great difficult in training and in generalization capabilities for large and complex problems.

Like NN, GA is a collection of algorithms based on the principles of evolution, heredity, and natural selection in living population and also deals with parallelism in data flow. GAs are adaptive, robust algorithms, and particularly useful for applications that require search and optimization, like providing efficient search and optimization tools for very large data sets, largely unconstrained by local minima problems. They don't require any mathematical modeling. Strictly referring to GA, its effectiveness consists of a high suitability for parallel computer implementation; particular success in large search and optimization problems; in an ability to learn complex relationships in incomplete data sets, their use as "data mining" tools for discovering previously unknown pattern; in their ability to provide an explanation of how decisions were produced, in a format that humans can understand; an ability to adapt to changes in their operating environment. GA limitations are as follows: they are computational expensive methods involving an application dependent setting of GA parameters (e.g. for crossover, mutation) that could be a time-consuming trial and error process. Typical successfully application to GA are many. The generation of control strategies for industrial processes (water treatment systems, cascaded chemical reactors, crackers, fermenters, distillatory columns, etc). There is the optimization of economic dispatch problem and unit commitment problem in electric power production, the use of optimized truck routing in delivery systems, optimization of manufacturing processes and job shop scheduling in industry. Aerospace applications use GA in the attitude determination of a spacecraft. There is the design of antenna and electronic circuits, including transducers. GA is also used in developing tailor-made solutions to individual customer requirements. And last, but not least, real-time adaptive hardware.

The above intelligent technologies have made great changes in various aspects of manufacturing industry. While the specific methodologies have led to encouraging results in several particular classes of problems, many complex tasks cannot be performed using a single technique alone. Hybrid systems combining the advantages of different, complementary paradigms are also important given the varied nature of application domains. Ongoing research in hybrid systems is attempting to address some of these problems.

Group Technology

Group technology (GT) is a manufacturing philosophy in which parts similar in geometry, material, or manufacturing attributes are identified and grouped together to take advantage of their similarities in manufacturing and design(Dowlatshahi & Nagaraj, 1998). It plays a major role in design standardization, manufacturing cell layouts, process planning, purchasing, and manufacturing technology systems design. The methodology of the application of group technology to manufacturing generally includes these steps: data collection, classification, coding, and analysis. And the main problem of GT is clustering and classification. GT is one of the key issues in the successful implementation of flexible manufacturing systems, and similarly is one of foundations of implementation of intelligent manufacturing.

The success of GT implementation is in the effective formation of part families and rational

layout of manufacturing cell (machine family). GT applications in CAD, process planning, and grouping scheduling are discussed in chapter 8. As the changes of manufacturing environment and the development of computer and network technology, GT should be further developed to adapt to these changes. Some new approaches (such as artificial neural network, fuzzy logic, and genetic algorithm, etc.) are used in GT, which provides part-machine formation with effective approaches. As an organizational approach which justifies small and medium batch sized production systems, GT is exploited in various phases of manufacturing and shows its advantages. GT will have to further improve so that it releases its new vigor.

MI VS. CONVENTIONAL TECHNOLOGIES IN MANUFACTURING

A manufacturing system can be conceptually thought of as being an integrated whole of complex interacting subsystems, organized in such a way as to synergically endeavor towards a common set of goals. The four main issues involved in manufacturing systems are design, planning, production, as well as diagnosis and maintenance.

Due to the inherent complexities associated with the manufacturing sub-systems, modeling them using common analytical and mathematical tools has proved to be very difficult, if not futile. Furthermore, conventional techniques of scientific computation and data processing are indispensable when automating each of these stages of manufacturing, there are a number of manufacturing problems for which these conventional techniques are ineffectual.

An intelligent technique can be employed in different contexts to execute different functional tasks, just as different techniques can be used to perform a particular task. The applications of MI to manufacturing bring great changes to traditional manufacturing systems. Therefore, it is necessary to discuss the advantages that caused by adoption of MI in the aforementioned four manufacturing main stages.

Design Process

In manufacturing systems, the customer needs are fed in, along with the feedback from the system and/or surroundings. The design process includes various tasks and individual tasks have different aspects depending on the products to be designed, the manufacturing environments of products, and the creativity needed in the design. Using the available technology related to materials, properties, etc., and to some extent human intuition, the product is conceived(conceptual design). Subsequently, the product is given a configuration and is parameterized. It is estimated that the choices made by the product designer in the preliminary design stage affect 60% of all product life cycle costs (Welch & Dixon, 1991). The choices include product function, behavior, form, structural complexity, and materials. The impact of preliminary design choices on manufacturing process and cost is usually significant.

Before the intelligent techniques adopted by manufacturing systems, products were mass-produced, without changes. The service life of products was relatively long. But individualization of the market has now become more important and; many more versions of a product are required. Complexity is a characteristic of today's products. The design of complex structures is normally performed by several different experts. Each expert may be responsible for a different task; each of these tasks may be further divided into subtasks. For example, structural engineers are normally jointly responsible for the preliminary design; the detailed structural design is accomplished by teams of structural engineers, detailers and draftsmen; and manufacturing requires a team of experts from different areas. In the recent past, several attempts have been made to develop an

integrated structural analysis and design system. Therefore rule-based, case-based, graph-based expert systems, artificial neural networks, GA, MAS are among the MI tools most widely utilized for manufacturing design.

Conventional programs consist of a set of statements whose order of execution is predetermined. These programs are very inflexible; updating requires considerable effort because the programmer has to locate the appropriate place to update in the predefined sequence. The programmer must ensure completeness, i.e. ensure the program performs correctly for a possible combination of conditions, and uniqueness of the solution, i.e. the output is unique for every possible input. The user imagines the program as a black box, where the program outputs the results for the input provided; the user does not have any idea what program has produced certain results. The expert system eliminates some of the impediments posed by conventional programs by making a clear distinction between the knowledge collected in knowledge bases and the control strategy. This partitioning allows for incremental addition of knowledge without manipulating the overall program structure; the programmer need not guarantee completeness. Furthermore, by ranking several alternatives either using an evaluation diagram or using inexact inference methods, several solutions can be provided for a particular set of input conditions, thus relaxing the uniqueness constraint.

From the viewpoint of the intellectual levels, the design tasks can be categorized into the three levels shown in: routine work, creative work, and intelligent work. In routine work, the structures of problems were clarified previously, and each problem can be solved routinely without referring to the experiential knowledge. Creative work, on the other hand, requires creative power, and can be carried out only by experienced designers. Creativity is one of the highest abilities of human beings, and the clue to clarifying the creative power has not yet been found. Intelligent work is considered as just between the routine work and the creative work, and has been performed by experienced designers. However, expert systems can be applied to the intelligent work, if the design process is well understood and structured. In the field of machine product design, expert systems have not yet been utilized for creative type design. The applications are limited to the variant type design and the modification type design. In these design processes, designers first refer to the database, which contains the information about the products previously designed and the relations among different products. Some candidate products are then selected for feasible design solutions. The retrieved candidates are modified and sometimes combined with other predesigned products in order to satisfy the design requirements. The role of expert systems in this process is to provide the designers with information about suitable design rules by examining the consistency of the design objects. The knowledge required in these processes is derived from examples of predesigned products.

As a conclusion, the utilization of MI in manufacturing design provides great facilities for the designers, counting on the decisions made by ES, improves the efficiency of design, as the concurrent design becoming a primary trend in modern enterprises, and reduces the cost. Furthermore, as virtual reality, rapid prototyping, visual simulation, develop at a high speed, the design will be become more intelligent.

Process Planning

Product design is made accessible to the process planning sub-system that transforms manufacturable design into its process plan with a schedule. The functions of process planning are to determine the sequence of machine tools, clamping devices, cutting tools, and so on from the product information represented by the product drawings and/or product model. The input to the process planning system is the geometric and technologi-

cal information about the parts designed by the CAD system, the assembly information, and the facility information. A number of researchers have employed MI approaches in dealing with various aspects and issues in process planning.

In order to speed up the process planning activities, databases are needed of possible processes, machining methods and tool combinations, from which a selection can be made after which a CAPP system automatically generates the NC programs. These CAPP systems in turn are often based on the use of process planning features that specify basic material processing patterns (e.g. bending, material removal, welding patterns). Fuzzy logic reasoning and hybrid techniques combining expert systems, neural network and fuzzy logic appear to be the preferred MI approaches in planning. The main role of MI in process planning is the implementation of: recognition of rough shapes of parts, determination of raw material, determination of preference relations among machining surfaces, recognition of important surfaces which have an essential function in the part, selection of a machining reference surface which has to be fixed on the tables of machine tools, selection of a suitable machining sequence, etc.

Currently, having more advanced CAPP systems available, process planning in many metal working factories is performed only a couple of days (and sometimes less) before actually processing a job. Consequently, it makes sense to take into account the actual work load on the shop floor when developing process plans for a new order, for instance with the aim to balance the load among various workstations. One way to significantly increase the loading flexibility on the shop floor is by developing several alternative job routings (Zijm, 1995).

In (Ming & Mak, 2000) the Kohonen self-organizing neural networks and Hopfield networks are adopted to solve problems in setup planning efficiently. Kohonen self-organizing neural networks are utilized, according to the nature of the different steps in setup planning, to generate setups in terms of the constraints of fixtures/jigs, approach directions, feature precedence relationships, and tolerance relationships. The operation sequence problem and the setup sequence problem are mapped onto the traveling salesman problem, and are solved by Hopfield neural networks.

The advantages of the MI application to process planning may be summarized as follows: the knowledge and inference process can be divided into modules; the global inference process can be controlled by representing the inference sequence on the blackboard. Based on the Digital Factory/Manufacturing concept, process planning system, production system and MI technologies are jointly used for optimizing manufacturing before starting production. In principle, by exploiting Digital Manufacturing, manufacturing enterprises are expected to achieve: shortened product development; early validation of manufacturing processes; faster production ramp up; faster time-to-market; reduced manufacturing costs; improved product quality; enhanced product knowledge dissemination; reduction in errors; increase in flexibility.

Production Control

With the process plan thus prepared, the product may be manufactured. The production system could be considered as completely autonomous or semiautonomous. In such cases, to ensure that the process is under control, it is necessary to monitor the process, obtain status information about the different process and product variables, diagnose the problems, and control the process so that the diagnosed problems can be rectified. The principal functionalities are: Process Modeling, Monitoring, Diagnosis, Control, Inspection, and Assembly. Knowledge based systems, neural networks and fuzzy logic, often combined in hybridsystems, are most frequently applied to the production functionalities mentioned above.

Production control is generally defined as the decision making and information transmission aimed at performing effective and economical

manufacturing by utilizing the manufacturing resources. Utilization of the computer is now indispensable to realizing production control for large manufacturing systems. Active research areas are summarized as follows: (i) Systematisation of data processing and optimization of overall production planning, and (ii) Dynamics of production control such as just-in-time method. In order to apply the production control systems to practical manufacturing systems, it is necessary to utilize specialized knowledge and expertise of experienced production managers and system engineers. They include the knowledge concerning the methods to solve the problems in production control, and the knowledge concerning the fundamental data processing and operations needed for solving the problems.

A significant amount of research work has been carried out in applying MI techniques to the various aspects, topics and issues in Production control. The KBS task is to take into consideration the following factors: estimated residual tool life; available spare tools; number of workpieces to be machined; and cost of tools, scraps, delays, etc. As a consequence, the KBS cell management can lower the cutting parameters; replace the tool; and reschedule the task, etc., with proper messages to other parts of the system. KBSs were developed for lifetime prediction of machine tool elements; integration of deep knowledge about production with shallow knowledge about fault-cause relationships; accommodation of unscheduled disruption, such as the breakdown of machines, high priority jobs, and power failures.

NN paradigms were employed for noise analysis of machine tools; identification of cutting conditions; identification of tool wear state; prediction of tool wear development; monitoring progressive tool wear in single point turning operations; estimating the remaining cutting tool life under given cutting conditions; tool wear sensing in quasi-orthogonal cutting; monitoring of manufacturing processes; monitoring of turning processes; developing thermal actuators to actively compensate for thermal deformation; developing a decision making process model for grinding operations with multi-stage structure consisting of feedforward and brain-state-in-a-box NNs; detecting the onset of chatter vibrations in grinding by removing the subjectivity of operator decision making.

The main advantages of the NN approach are: no need to know a previous model to relate temperatures and deformations; the NN can learn linear and nonlinear models; most of the operations can be automated and the quality of results is independent of operator skill. Some disadvantages are present: it is not easy to program a NN in a simple PLC, which may imply the need of an external computer.

FL decision support systems were used for the evaluation and selection of cutting conditions; monitoring work material heat treatment conditions during machining; pattern recognition of metal cutting states; control of machine tools to obtain the desired surface finish by controlling the machining process; nondestructive inspection of composite material electrical insulators.

Diagnosis and Maintenance

Despite the quality of a product being mostly dependent on intelligent facilities, even if a condition monitoring program is in operation, any negligible error or disorder in the process, can still cause a catastrophic system failure, defeating the very purpose for which the investment was made in condition based maintenance. Moreover, any unexpected process disorder can decrease production capability and the product quality. To avoid such an unexpected process disorder, effective maintenance is necessary to increase the reliability of the process by diagnosis and manitenance. For the effective maintenance, the fault diagnosis is preceded before causing a catastrophic system failure. The principal functions of diagnostic systems are as follows: The system first detects the indications of abnormal state of the objects,

and then estimates the causes of the abnormal state. Finally, suitable counterplans for the estimated causes are determined (Moubray, 1997). MI technologies can be applied to each of these three functions and many MI technologies, like expert systems, artificial neural networks, fuzzy logic, etc, for diagnosis have been developed and put into practice.

During the diagnostics process, specific signals are typically compared to threshold limits. Additional processing may determine a signature pattern in one, or multiple, fault measure(s). Automated reasoning is often used to identify the faulty type (cracked gear tooth, bearing spall, imbalance etc.), location, and severity. The core problem of diagnostics is essentially a problem of classification. Discriminant transformations are often used to map the data characteristic of different failure mode effects into distinct regions in the feature subspace. A common inference engine can be applied to the diagnosis of different machines by modifying the rules. The results of diagnosis can be applied to the maintenance and operation of machines. That is, a total system for diagnosis, maintenance and operation can be formed.

Because the limitation of search space in a diagnosis problem, it is possible to explain the inference process of diagnosis, deal with ambiguous relations. And the inference process is adaptable to the diagnosis of various types of machines; as a result, MI is applied to diagnostic systems successfully.

In recent years, various strategies have been reported for the diagnosis of manufacturing systems. The intelligent systems used in condition-based fault diagnosis can be divided into three categories such as rule-based diagnostic systems, case-based diagnostic systems, and model-based diagnostic systems (Milne, 1989). Rule-based diagnostic systems detect and identify incipient faults in accordance with the rules representing the relation of each possible fault with the actual monitored equipment condition. Case-based diagnostic systems use historical records of maintenance cases to provide an interpretation for the actual monitored conditions of the equipment. A model-based diagnostic systems use different mathematical, neural network, and logical methods to improve diagnostic reasoning based on the structure and properties of the equipment system. Among above three diagnostic systems, the rule-based systems usually consist of a bunch of IF-THEN rules, a bunch of facts, and some interpreter controlling the application of the rules, given the facts. Due to this structure that the humans can understand them easily, the analysis such as forward or backwards reasoning is easier than the other diagnostic systems. The current process condition can be inferred by forward reasoning with the current values of the attributes, and the possible maintenance actions can be figured out by backward reasoning. ESs have been realized for conceptual diagnosis and maintenance of FMS working stations; keeping the knowledge on fault behaviour of automated assembly systems and making it available for diagnosis in the shop floor; tool condition assessment in order to avoid unnecessary interference with machining process.

Although NN has proven to be a powerful technique for detecting and classifying fault types, it is regarded as a 'blackbox' technology because of the lack of comprehensibility. The knowledge learned by NNs is hard to be understood for users, because it is concealed in a large amount of connections. This is a significant shortcoming, for without the ability to produce understandable decisions. Therefore, although NNs are capable of learning the complex nonlinear relationships between process parameters and fault types in manufacturing processes, they cannot explicitly provide rules to pinpoint the causes of these faults. To solve the 'blackbox' problem of NNs, many researchers have addressed the issue of improving the comprehensibility of NNs. Although these ANN-based rule extraction approaches can provide the domain theory, these algorithms, in general, are so complex to result in too much com-

putational time. Moreover, those rules extracted from specific NNs are often incomprehensible to users, which limits their effective utilization. Thus, these drawbacks of NN based rule extraction approaches result in potentially a hindrance to real-time process monitoring and control.

In recent years, there have been numerous attempts to apply evolutionary algorithms (EAs) in data mining to tackle the problem of knowledge extraction and classification. Among the different areas in knowledge discovery, GAs have achieved vast popularity in the application of rule induction, since the sets of IF-THEN rules of classification problems can be easily represented by choosing an encoding of rules that allocate specific substrings for each rule pre-condition and post-condition. A GA-based rule extraction algorithm to extract rules from the manufacturing process, which are used to express the causal relationships between process parameters and product output measures. Moreover, these rules are used for constructing a KBANN(Knowledge Based Artificial Neural Network) with high performance for monitoring abnormal signals in manufacturing systems. With a seamless integration of the KBANN and the GARule approach software, the proposed system not only pinpoints the quality problems of the manufacturing process (by the KBANN), but also reveals the causal relationships of why they occur and how to prevent them (by the GARule). The successful implementation of the developed methodology in an industrial case demonstrates its effectiveness and potential applicability for these manufacturing processes, where there existed the causal relationships between the process parameters and product output measures.

Agent technology is emerging as better decision support tools than traditional Artificial Intelligence. Therefore a lot of diagnosis strategies are proposed, which adopt agent or MA technology.

The concept of remote monitoring, diagnosis and maintenance for globally integrated manufacturing facilities and activities was first introduced by (Lee,1997). The author proposed six functional requirements for the purpose. They are: (i) Multi-Sensor Integrated Monitoring And Control Systems: multi-sensors are required to keep track of the performance and condition of the machines remotely. A watchdog agent, which might use neural networks, may also be needed to provide on-line composition and reasoning. (ii) Communication and integration of geographically dispersed machines through a multimedia information environment. (iii) Data abstraction: only the relevant data are to be transmitted through the network. (iv) Knowledge acquisition and learning: there is a need for intelligent tools to acquire and organize data on the manufacturing processes at one site to share it with other sites. The whole system should be capable of learning the behaviour of users at different sites. (v) Natural language translation: tools to automatically translate text into other languages. (vi) Tele-Maintenance and Collaborative Diagnostics to facilitate for the technical personnel to perform diagnostics on machines that are geographically distributed. An Intelligent Control Maintenance Management System (ICMMS) with the objective of analyzing the equipment of an automation system based on a multi-agent system approach is proposed in (Fu, et al, 2002). It has been applied to a hydroelectric generating unit. The ICMMS architecture consists of several agent blocks such as the perception block, the monitoring and diagnosing block, the planning block, the intention block, the actions block and the knowledge base. All the agents have a similar structure. The only thing that varies is their behaviour and knowledge. The agents accomplish different tasks and cooperate with each other to achieve a common goal. A MAS developed on the basis of a reference model for fault management systems is proposed in (Cerrada, et al, 2007). They use fault management systems to develop a model to perform preventive maintenance plans and detection, isolation and diagnosis tasks. It takes into account detection, fault isolation and diagnosis, fault cause identification, preventive

and corrective maintenance tasks, maintenance task planning and fault prediction model and development. This is attained by the coordinated interaction of the agents.

Developments in the application of MI technologies to diagnosis and maintenance are at a rudimentary stage. Only a few consistent and systematic efforts in this direction have been found. The complete integration of data, systems and processes emerges as an important requirement to fulfil. Successful implementation of diagnosis and maintenance, especially when the assets are variously located geographically, is becoming a key issue of enterprises. What should be expected in the future is that the use of Web technologies in conjunction with MI technologies will increase in use for diagnosis and maintenance integration purposes. Moreover, these technologies can also integrate different processes and diagnosis and maintenance systems.

PROSPECT OF MANUFACTURING INTELLIGENCE

The manufacturing system has been in the process of constant perfection and development in order to meet the needs of the changing market. Intelligence is the extension of the manufacturing system on the basis of flexibility and integration (Beach, 2000). As the manufacturing system changes from an energy driven mode to an information driven mode, intelligence and flexibility are necessary features for the manufacturing system (Anon, 2007) in order to cope with the great amount of complicated information, changing market demand, fierce competition and complicated environment. Manufacturing Intelligence (MI) studies how to adopt artificial intelligence and computing intelligence methods to solve problems in digital manufacturing, such as intelligent digital scheduling, intelligent digital designing, intelligent digital machining, intelligent digital controlling, intelligent digital process planning, intelligent digital maintenance and diagnosis. This definition shows that the development of manufacturing intelligence technology is critical to the future of digital manufacturing. Before discussion of the development trend of manufacturing intelligence, the key theoretical problems of manufacture intelligence are considered.

1. **Gradual complication of the manufacturing system.** In order to be sensitive to the fast response and restructure of the manufacturing system, the research results of many subjects such as information science and life science and sociology have to be considered and used. To exploit the new structure system, the manufacturing mode and effective operation mechanism, an optimized organization structure and good operation, form the main object of modeling and optimization. The new structure system of manufacturing is not only of great importance to the enterprise's sensitivity and response ability to requirement changes and restructure, but also to the increased requirement for flexibility of manufacturing devices and ability to dynamically restructure.
2. **More complex basic computing.** In CAD/CAM integration, CMM and robotics, in 3-real space, there exist many design and analytical problems of geometrical algorithm, especially geometrical expression, geometrical computing and geometrical reasoning; during measuring and planning a robot's way and localization, there exists geometrical computing and geometrical reasoning of the configuration space. During description of robot's hands catching and getting plan, assembling plan and operation, there exists geometrical reasoning of screw space. During the process of manufacturing, geometrical research of physical and mechanical phenomenon has brought many research subjects, like geometrical computing and geometrical reasoning in

manufacturing science. And it leaves too much to be desired, so far, one new subject –computer geometry is attracting more and more attention.

3. **Greater complexity in manufacturing information.** Enhancing the ability to process manufacturing system information has become a key feature of modern manufacturing science (Heragu, 2002). Gaining and integration of manufacturing information become tridimensional. And a multi-layer information organization and structure is appropriated. The structure modeling of manufacturing information, consistency of manufacturing information, transmission process and knowledge based management, still leaves much to be desired.
4. **Emergence of various intelligent technologies.** An intelligent computing tool based on the evolution algorithm, in the field of organization optimization solution technology including operation problems, is receiving wide attention, and there is optimism that it will deal with the problems of solution speed and solution accuracy. This is only one example; manufacturing intelligence also embodies intelligent operation, intelligent design, intelligent process, robotics, intelligent control and intelligent technique planning and intelligent detection.

These are all key theoretical problems for the development and creation of an intelligent manufacturing system, and their solution will ensure that manufacturing updates from one technique to a science.

The development path of technology is frequently beyond human imagination, and it is therefore almost impossible to predict the future of intelligence technology. Based on some advanced research work in this field, however, it is most likely that manufacturing intelligence technology will evolve in the directions outlined below.

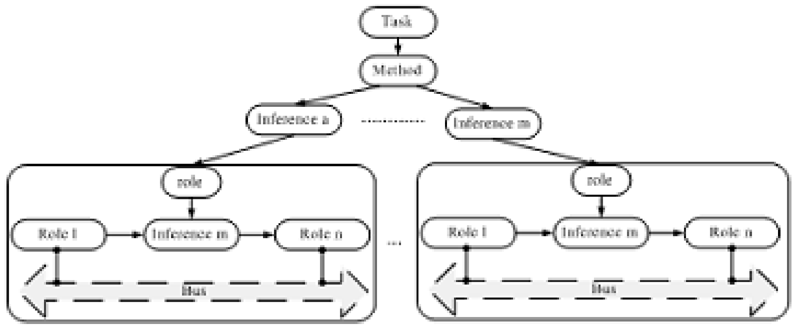
Update Theoretical Frame

As the core of the intelligent manufacturing system, the key to intelligent technology (such as intelligent computing, and intelligent robot) is artificial intelligence (AI) (Devedzic, 1999). Their interaction will bring about an intelligence 'nuclear explosion' and stimulate man's potential intelligence, finally resulting in an intelligent revolution, which will develop the manufacturing industry in a creative and dynamic explosive way.

Comparing short-term and long-term goals, however, there are weak points in the research of human intelligence, which mainly feature the following aspects:

1. **Great distance between micro and macro.** On one hand, philosophical, perceiving science, thought science and psychology are too high-leveled and abstract; on the other hand, human intelligence logic symbolization(as shown in Figure 3), neural network and behaviorism is too low-leveled. There is a great gap between the two, and some research points between them have not been given attention, making it hard to connect them effectively. Nowadays, it is widely accepted that the general purpose of AI is to develop conceptual models, formally rewriting processes of these models, and programming strategies and physical machines to reproduce as efficiently and thoroughly as possible the most authentic cognitive, scientific and technical tasks of biological systems that have been labeled intelligent. The scientific perspective of AI looks for a computable theory of human knowledge, in other words, a theory in which its formal models can be executed in a calculus system and have the same predictive character that, for example, Maxwell's equations have in electromagnetism. It is therefore not surprising that with these objectives the strong conjecture of AI

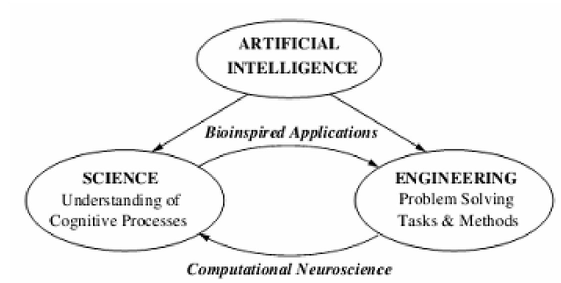
Figure 3. Symbolic modeling approach



excessive is considered. Giving biology and psychology the character of experimental science that physics has is a desirable objective but it is difficult to achieve.

2. **Distance between integration and localization.** Human intelligence is the whole reflection of the brain system, which contains multiple layers and many aspects. Symbolization only emphasizes the abstract thought characteristics of the human mind, and links simply imitate visualized thought characteristics; behaviorism, however, cares about behavior characteristics and the evolution of human intelligence. All have obvious weak points, so it is necessary for research on them to be carried out from a multi-layered, multi-factorial and multi-dimensional position.
3. **Disjunction of theory and practice.** The general working of the human brain is understood, but it is difficult to get a deep understanding of intelligence since it changes so fast and is full of variety. In its microcosmic world, little is known about its working rules and all the intelligent theories put forward appear to be so much wishful thinking, so it's quite successful to show 'intelligent' in some aspects.(as shown in Figure 4). Because of technological limitation, this disjunction phenomenon will last for a long time. The many problems about

Figure 4. Relationship between AI theory and engineering



intelligence and similar questions illustrate that human brain structure and functions are highly complicated, as are related tasks. If a complete understanding of the brain's structure and functions, problem-solving abilities is to be discovered, and successful research tasks of artificial intelligence are to be undertaken, a newer AI frame has to be discovered and set up, and theoretical structure created to lay a theoretical foundation for its further development. The research in the biology information field and genetics can hopefully provide new thought and results for AI theoretical research.

So far, the reasoning function of AI has progressed far, and its study and imaginative functions are the subject of research, and the next step of research is to imitate roughness function of right

head and parallel processing function. Artificial neural networks will be the new field of future AI applications; future intelligent computers will be a combination of host and artificial neural networks. According to research, emotion is one part of intelligence which cannot be separated from the whole of intelligence. so the next stage of AI research will realize human emotion in the computer. Emotion is very important for natural communication between computers and humans, and current AI research is developing towards numeric-based computer intelligence.

Greater Technological Integration

Manufacturing intelligence is an integration of other information process technologies and related subject technologies. Realizing this integration presents many challenges, for instance, creating a standard mode of knowledge representation and transmission, understanding effective interaction among sub-systems and developing new methods of knowledge of complex expressions of numeric mode and non-numeric mode, also including the combination of quantity fixation, to reason naturally at a fast speed. Besides digital technology, information technology that needs to be integrated includes computer networks, remote communication, database, computer, sound and video, robotics, process control, parallel computing, light computing and biological information processes. In recent years, Agent technology and digital exploit have also provided wide support for the technological integration of manufacture intelligence (Marik, 2005; Pechoucek, 2007). Manufacturing intelligence systems not only contain modern computing technology, perceiving science and computer languages, but also manufacturing science and system science.

Computing is not only an important part of the manufacturing system support structure, it is also the energy of system, as blood is to the human body. New computing methods and ways to realize intelligence, such as FL, NN, GA, and Chaos, are all part of computing intelligence (Aytug, 2003). ES is better for improving the intelligence level. Nowadays, the research application emphasis of computing intelligence lies in FL, NN and GA, and in the IM research field, there are: intelligent transmitter, intelligent control during process, intelligent detection and monitor in manufacturing system, cutting intelligent optimization, machinery components stability analysis and optimization design, machine error intelligent detection, intelligent study and decision and forecast (Tan, 2006). The development of an intelligent manufacturing system, cognitive science and knowledge technology science, will undoubtedly cause an enormous impact on future industry and society.

Many research domains are relevant in bringing the manufacturing systems of the future to success. There is abundant know-how available in the scientific world; the challenge is to integrate it into a consistent, autonomous frame. Autonomy is one of the most important factors for intelligent manufacturing systems of the future, as it allows systems to reach much higher levels of service to the user.

Numerous manufacturing companies report that it is becoming more difficult to recruit skilled staff to operate production machines; it seems that this kind of profession is no longer popular. In some cases, off-shoring to other countries or continents may be a solution (with the disadvantage of potential loss of business), but particularly in domains that require a high level of precision, automation is the only alternative. The more autonomous systems are, the better.

In countries with high salaries, it is too expensive to employ staff to supervise the machines at night. This means that the machines run as long as everything goes well, and in the case of minor problems, they simply stop production. Valuable production time is lost every night. Some system autonomy to handle complications could avoid this.

If systems can look after themselves to a substantial degree and can propose solutions to the types of problem which frequently occur, there is less need for highly specialized staff. Selecting from limited proposals as to how to proceed is much easier than having to imagine what to do from zero. In this sense, limited autonomy is a way of assisting the user as much (or as little) as he or she desires.

The more complex the system, the more assistance the user may require. High complexity in the body needs corresponding complexity in the brain; this means that the human user can easily become overburdened by managing systems with many modules, many products and many interactions.

A system which is made to run autonomously also has to be able to cope with changes, be it the addition or removal of modules or being confronted with new requirements without programming. In other words, autonomy leads to evolvement.

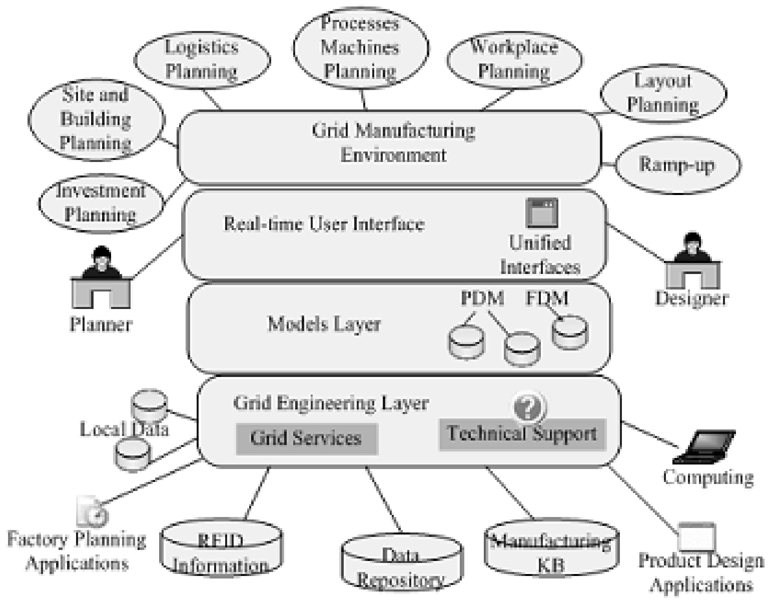
Mature Application Method

The key of manufacturing enterprise intelligence is to use IMS(Intelligent Manufacturing System). In the beginning, intelligent manufacturing systems used expert systems to enhance manufacturing intelligence, such as machine tool self-adaptability control, which intelligent behavior embodies at symbolization reasoning. Most of these ES are not real time systems; data are not updated and there is no information communication with the outside, so it is a low-level closed intelligent system. In order to avoid the defects like, ES dependence on experts, difficulties in obtaining knowledge, inconvenience of expressing in logic knowledge and unpopularity, the modern computer, according to the Van Neumann principle, combines the Van Neumann host with a human neural network as outside intelligence to form an intelligent machine. With the development of the computer towards being an intelligent machine, CIMS(Computer Integrated Manufacturing System) will undoubtedly develop into CIIMS(Computer Integrated Intelligent Manufacturing System), and become a true intelligent manufacturing system.

The realization of an intelligent manufacturing system needs guaranteed quality of hardware, but software is the key technology of intelligent manufacturing. Many intelligent manufacturing systems need to develop complicated software systems, which is beneficial for the development of software engineering. Software engineering can provide standard programs for solving some types of problems. Knowledge software provides effective programming tools for intelligent manufacturing problems, but because of the complication and popularity of its problems, traditional software design methods are not sufficiently capable nor suitable. The functions that intelligent manufacturing software has to perform are likely to change in accordance with system development. A manufacturing intelligent method has to support its developing experiments and allow the system to develop a complete application system from a smaller core model in an organized way, so it becomes urgent for the enterprise to utilize all the resources within or outside it to satisfy the above requirements, i.e., an intelligent manufacturing system based on the web still needs to be developed in a way that integrates visualization based systems and remote application services to facilitate product design and manufacturing on the web. Manufacturing Grid (MG) technology is the potential solution to this situation.

MG is an integrated supporting environment both for the sharing and integration of resources in enterprise and social and for the cooperating operation and management of the enterprises. Based on the grid and relatively advanced computer and information technologies, MG shields the heterogeneousness and the regional distribution of resources by way of encapsulating and integrating the design, manufacture, management, information, technology, intelligence and software resources found separately in different enterprises and social groups. It not only provides

Figure 5. Structure of manufacturing grid



various manufacturing services for customers in a transparent way, but also makes enterprises or individuals conveniently obtain all the services related to manufacturing by way of requesting services, and to use all the resources encapsulated in the manufacturing grid with as much convenience as using the local ones. It achieves the integration and optimal operation of all kind of resources and provides a cooperative work environment for the construction of the manufacturing grid application system faced with the special requirements of network manufacturing.

MG is a manufacturing-oriented virtual network on the basis of Internet, Grid, and other related technologies, which aims to enable the collaborative sharing of resources and competencies in manufacturing engineering and which supports the integrated and continuous management of the entire product and factory life cycle. MG provides manufacturing enterprises and individuals with manufacturing services in a similar way that the Internet provides information services. Furthermore, MG supports the collaborative planning, operation and management of manufacturing in close relation with product design and development and responds to emerging challenges with innovation, speed and flexibility.

As shown in Figure 5, the MG is structured in three layers, in keeping with current research trends. The first layer, called the Grid Engineering Layer, networks by employing Grid computing the distributed product design and manufacturing planning resources and activities that are interconnected for fulfilling the Factory Planning Life Cycle phases. These are represented mainly by the specific applications for product design and factory planning, for example, ProEngineer (PTC), Tecnomatics components (Siemens PLM Software) and RFID applications. The Grid services comprise all the activities which have to be fulfilled, such as product design, product assessment, factory layout and manufacturing execution, which access and use data from their corresponding databases. The RFID information is used for synchronizing the digital factory, for example, with context and situation-aware data of production resources and tools coming from the real production on the shop-floor.

The Models Layer is the layer which integrates the Factory Data Model, the Product Data Model and the reusable simulation models already developed for supporting the product design and factory planning activities, for example, the material flow simulation, the acoustic simulation, work-place and collision detection simulation. These models are integrated with the corresponding Product and Factory Data Models in order to access and use the required data.

The On-line Real-time User Interface layer enables the user to access the design and planning environment through the web portal, to accept requests on services and to display results, respectively the online monitoring. The conceptual work and prototyping of the environment is ongoing, facing the challenge of selecting the suitable Grid technology for realizing the Grid Engineering Layer. The expected benefits generated by implementing the presented approach are as follows: a) continuous design of products and planning of factory by orchestrating the simulation activities in both worlds, product and factory, and b) feedback from the later stages can be communicated back to the design and planning stages in order to reduce future implementation errors and the required time for product launching and factory ramp-up.

A regional MG system can be built in special region, and by combining different regional MGs, a whole MG system can be constructed to support network manufacturing. Based on MG, future enterprises, and even individuals, could obtain various manufacturing services from the Internet as conveniently as daily obtaining water, electricity, and gas.

Apart from automation, novel processes, materials, MI technologies and new concepts in manufacturing have made their appearance and are anticipated to change the way manufacturing is organized today. Several topics remain open and unanswered in MI and could be researched in the future as well as new and promising research windows will open related to this research domain. This section intends to give a bird's-eye view of trends and challenges in MI technology. However, MI technologies cannot solve all challenges introduced by the requirements of mobility, modularity and re-configurability, etc. In this way, MI technologies must be integrated with other new technologies, such as web-based technologies (including web services Semantic Web, and wireless networks), embedded systems, RFID systems and grid computing for its wide and successful applications in industry in a near future.

SUMMARY

In this chapter, the development trends and prospects of MI have been discussed on the basis of analysis of the various characteristics of the manufacturing industry in the 21st century. The general trends in research and development of MI and their roles in manufacturing are summarized. Then, a brief comparison of MI technologies and conventional technologies in manufacturing is discussed taking some representative systems for various manufacturing fields, such as design, process planning, production control and diagnosis. Finally, the future directions of MI technologies are considered. Global manufacturing systems are today on the road toward digitalization. The development of the manufacturing system needs to incorporate an intelligent system with features such as self-learning, self-adaption, self-organization and self-maintenance; to this end, digitalization is the key, intelligence is the prospect, networking is the focus, and cleaning is the direction. Manufacturing technology will develop towards precision, extreme, high speed and standardization. These aspects have complicated interactions and interrelationships, and dynamic changes as a whole.

During the process of writing this book, various opinions about the work were obtained, including expert opinions, and an optimistic attitude towards the future of manufacturing intelligence is held. It is pleasing to see that many research results of

manufacturing intelligence have already entered and affected people's daily lives, for example, in the application of the powerful tools in engineering, produced by intelligence technology, which have become more widespread due to the power and affordability of present-day computers. Other technological developments in manufacturing intelligence that will have an impact in engineering include data mining, or the extraction of information and knowledge from large databases, and multi-agent systems, or distributed self-organizing systems employing entities that function autonomously in an unpredictable environment concurrently with other entities and processes. Future agriculture production will be changed from self-control to intelligent control, realized by the intelligence of the whole process. 'Plant factory' has originated in Japan, which predicts a great future for the agricultural factory, with automation, and even intelligence. Intelligent machines are not only seen in factories and farmlands, but also in offices and the home; 'intelligent building' and 'intelligent house' are being built which will feature computer control, robot service and network communication, making intelligent offices and houses popular, and finally realizing intelligent cities. Apart from industry, business, health-care and national defense, manufacturing intelligent technology has been widely used in the fields of transportation, agriculture, air, communication, culture, education, management and decision-making, and information search.

It is certain that manufacturing intelligence will have a bright future despite the fact that there is still a long way to go. Meanwhile, the high cost of research and the hard work of many generations will continue. The development of any science or technology is not easy, and the future development of manufacturing intelligence may encounter new difficulties, even great setbacks, and probably risk faced by researchers. The appropriate deployment of manufacturing intelligence will contribute to the basic research and creation of more competitive engineering systems.

REFERENCES

Anon., (2007). Technologies that could change manufacturing. *Manufacturing Engineering*, *139*(6), 21–22.

Aytug, H., Khouja, M., & Vergara, F. E. (2003). Use of genetic algorithms to solve production and operations management problems: a review. *International Journal of Production Research*, *41*(17), 3955–4009. doi:10.1080/00207540310001626319

Beach, R. (2000). A review of manufacturing flexibility. *European Journal of Operational Research*, *122*(1), 41–57. doi:10.1016/S0377-2217(99)00062-4

Cerrada, M. (2007). Agent based design for faultmanagement systems in industrial processes. *Computers in Industry*, 313–328. doi:10.1016/j.compind.2006.07.008

Cheng, K., Pan, P. Y., & Harrison, D. K. (2001). Web-based design and manufacturing support systems: implementation perspectives. *International Journal of Computer Integrated Manufacturing*, *14*(1), 14–27. doi:10.1080/09511920150214875

Chisholm, A. W. J. (1990). Nomenclature and definitions for manufacturing systems. *Annals of the CIRPM*, *39*(2), 735–744.

Devedzic, V., & Radovic, D. (1999). A framework for building intelligent manufacturing systems. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *29*(3), 422–439. doi:10.1109/5326.777077

Dimopoulos, C., & Zalzala, A. M. S. (2000). Recent developments in evolutionary computation for manufacturing optimization: problems, solutions, and comparisons. *IEEE Transactions on Evolutionary Computation*, *4*(2), 93–113. doi:10.1109/4235.850651

Dowlatshahi, S., & Nagaraj, M. (1998). Application of Group Technology for Design Data Management. *Computers in Engineering*, *34*(1), 235–255.

Duvivier, D. (2003). Integrating simulation and optimization of manufacturing systems. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *33*(2), 186–192.

Fu, C., et al. (2002). MAS-based model of intelligent controlmaintenance-management system (ICMMS) and its application. In *Proceedings of 2002 International Conference on Machine Learning and Cybernetics*, (pp. 376-380).

Halevi, G., & Wang, K. (2007). Knowledge based manufacturing system (KBMS). *Journal of Intelligent Manufacturing*, *18*, 467–474. doi:10.1007/s10845-007-0049-1

Heragu, S. S. (2002). Intelligent agent based framework for manufacturing systems control. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, *32*(5), 560–573. doi:10.1109/TSMCA.2002.804788

Kahn, K. B. (2001). Market orientation, interdepartmental integration, and product development performance. *Journal of Product Innovation Management*, *18*(5), 314–323. doi:10.1016/S0737-6782(01)00101-1

King, A. A., & Lenox, M. J. (2001). Lean and green? An empirical examination of the relationship between lean production and environmental performance. *Production and Operations Management*, *10*(3), 244–256.

Koh, S. C. L., & Gunasekaran, A. (2006). A knowledge management approach for managing uncertainty in manufacturing. *Industrial Management & Data Systems*, *106*(3-4), 439–459. doi:10.1108/02635570610661561

Kwon, Y., & Rauniar, S. (2007). E-quality for manufacturing (EQM) within the framework of internet-based systems. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *37*(6), 1365–1372. doi:10.1109/TSMCC.2007.905827

Lastra, J. L. M., & Delamer, M. (2006). Semantic web services in factory automation: fundamental insights and research roadmap. *IEEE Transactions on Industrial Informatics*, *2*(1), 1–11. doi:10.1109/TII.2005.862144

Lee, J. (1997). Strategy and challenges on remote diagnostics and maintenance for manufacturing equipment. *Proceedings of the Annual Reliability and Maintainability Symposium*, (pp. 368-370).

Li, P. G. (1998). Manufacturing system performance ananlysis modeling—Theory and Method. Wuhan, PRC: Huazhong University of Science and Technology Press.

Mabert, V. A., Soni, A., & Venkataramanan, M. (2003). Enterprise resource planning: Managing the implementation process. *European Journal of Operational Research*, *146*(2), 302–314. doi:10.1016/S0377-2217(02)00551-9

Madhusudan, T. (2005). An agent-based approach for coordinating product design workflows. *Computers in Industry*, *56*(3), 235–259. doi:10.1016/j.compind.2004.12.003

Marik, V., & McFarlane, D. (2005). Industrial adoption of agent-based technologies. *Intelligent Systems*, *20*(1), 27–35. doi:10.1109/MIS.2005.11

Mehrabi, M. G., Ulsoy, A. G., & Koren, Y. (2000). Reconfigurable manufacturing systems: Key to future manufacturing. *Journal of Intelligent Manufacturing*, *11*(4), 403–419. doi:10.1023/A:1008930403506

Milne, R. (1989). Strategies for diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics*, *17*(3), 333–339. doi:10.1109/TSMC.1987.4309050

Ming, X. G., & Mak, K. L. (2000). Intelligent setup planning in manufacturing by neural networks based approach. *Journal of Intelligent Manufacturing*, 311–331. doi:10.1023/A:1008975426914

Mircea, G., et al. (2005). Computational Intelligence: Engineering of Hybrid Systems, 174, 1-11. Berlin: Springer-Verlag.

Mok, S. M. (2001). Modeling automatic assembly and disassembly operations for virtual manufacturing. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans*, *31*(3), 223–232. doi:10.1109/3468.925662

Monostori, L. (2006). Agent-Based Systems for Manufacturing. *Annals of the CIRP*, *55*(2), 697–720. doi:10.1016/j.cirp.2006.10.004

Moubray, J. (1997). *Reliability-Centered Maintenance* (2nd ed.). New York: Industrial Press.

Paul, K. W. (2002). *21 Century Manufacturing*. New York: Prentice Hall.

Pechoucek, M. (2007). Agent-based approach to mass-oriented production planning: case study. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *37*(3), 386–395. doi:10.1109/TSMCC.2006.889537

Pham, D. T., & Pham, P. T. N. (1988). Expert Systems in Mechanical and Manufacturing Engineering. *International Journal of Advanced Manufacturing Technology*, *3*(3), 3–21. doi:10.1007/BF02601587

Rosenzweig, E. D., Roth, A. V., & Dean, J. W. (2003). The influence of an integration strategy on competitive capabilities and business performance: An exploratory study of consumer products manufacturers. *Journal of Operations Management*, *21*(4), 437–456. doi:10.1016/S0272-6963(03)00037-8

Shah, R., & Ward, P. T. (2003). Lean manufacturing: context, practice bundles, and performance. *Journal of Operations Management*, *21*(2), 129–149. doi:10.1016/S0272-6963(02)00108-0

Sun, J., Zhang, Y. F., & Nee, A. Y. C. (2001). A distributed multi-agent environment for product design and manufacturing planning. *International Journal of Production Research*, *39*(4), 625–645. doi:10.1080/00207540010004340

Tan, K. H. (2006). An intelligent decision support system for manufacturing technology investments. *International Journal of Production Economics*, *104*(1), 179–190. doi:10.1016/j.ijpe.2005.02.010

Wang, G., Huang, S. H., & Dismukes, J. P. (2004). Product-driven supply chain selection using integrated multi-criteria decision-making methodology. *International Journal of Production Economics*, *91*(1), 1–15. doi:10.1016/S0925-5273(03)00221-4

Wang, H. H., Zhou, Z. D., & Chen, Y. P. (2006). A methodology for evaluating the manufacturing ability of equipment in virtual enterprises. *International Journal of Computer Integrated Manufacturing*, *19*(4), 339–349. doi:10.1080/09511920500064714

Ward, P. T., & Duray, R. (2000). Manufacturing strategy in context: environment, competitive strategy and manufacturing strategy. *Journal of Operations Management*, *18*(2), 123–138. doi:10.1016/S0272-6963(99)00021-2

Wayne, C. T. (2002). *Introduction to Industrial and Systems Engineering* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Welch, R., & Dixon, J. (1991). Conceptual design of mechanical systems. In *Proceedings of ASME Design Theory and Methodology Conference*, (pp. 61-68).

Yoshimura, M. (2007). System design optimization for product manufacturing. *Concurrent Engineering . Research and Applications*, *15*(4), 329–343.

Yuan, X. B., & Yang, S. X. (2003). Virtual assembly with biologically inspired intelligence. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *33*(2), 159–167. doi:10.1109/TSMCC.2003.813148

Zhou, Z. D., & Liu, Q. (2006). The connotation of digital manufacturing science and its expectation. *1st International Symposium on Digital Manufacture, 3*(1), 1-8.

Zijm, W. H. M. (1995). The integration of process planning and shop floor scheduling in small batch part manufacturing. *Annals of the CIRP*, *44*(1), 429–432. doi:10.1016/S0007-8506(07)62357-9

About the Authors

Zude Zhou is a professor at the School of Mechanical and Electronic Engineering, Wuhan University of Technlogy, P. R. China. He is also the President of Wuhan University. He received his Bacholar degree in Electrical Engineering from Huazhong Univeristy of Technolgy, China, in 1970. He attended in a advanced studies at Birmingham,UK, from 1984-1986. He was also a Visiting Professor of the University of Bolton,UK, a Visiting Professor of Hongkong University, Hongkong, and a Visiting Scientist of National University of Singapore, Singapore. Mr. Zhou specializes in control and application of microcomputer, electromechanical integration technology, intelligent manufacturing, digital manufacturing, network manufacturing. He has published more than 300 articles.

Huaiqing Wang is a Professor at the Department of Information Systems, City University of Hong Kong. He is also the Honorary Dean and a Guest Professor of the School of Information Engineering, Wuhan University of Technology, China. He received his PhD in Computer Science from University of Manchester, UK, in 1987. Dr. Wang specializes in research and development of business intelligence and intelligent agents applications, such as financial intelligence, manufacture intelligence, knowledge management systems, and ontology. He has published more than 60 SCI/SSCI journal articles, including *Communications of the ACM, Artificial Intelligence, and many IEEE Transactions.*

Ping Lou is an Associate Professor at the School of Automation, Wuhan University of Technology, P. R. China. She received his PhD in Mechanical Engineering from Huazhong University of Science and Technology , P. R. China, in 2004. Dr. Lou specializes in research and development of intelligent manufacturing and network manufacturing. She has published more than 20 articles.

Index

C

D

J

K

L

M

N

O

P