Preface

Today's business can be described by a single word: turbulence. Turbulent markets have the following characteristics: shorter product life cycles, uncertain product types, and fluctuating production volumes (sometimes mass, sometimes batch, and sometimes very small volumes).

In order to survive and thrive in such a volatile business environment, a number of approaches have been developed to aid companies in their management decisions and engineering designs. Among various methods, data mining is a relatively new approach that has attracted a lot of attention from business managers, engineers and academic researchers. Data mining has been chosen as one of ten emerging technologies that will change the world by *MIT Technology Review*.

Data mining is a process of discovering valuable information from observational data sets, which is an interdisciplinary field bringing together techniques from databases, machine learning, optimization theory, statistics, pattern recognition, and visualization.

Data mining has been widely used in various areas such as business, medicine, science, and engineering. Many books have been published to introduce data-mining concepts, implementation procedures and application cases. Unfortunately, very few publications interpret data-mining applications from both management and engineering perspectives.

This book introduces data-mining applications in the areas of management and industrial engineering. This book consists of the following: Chapters 1–6 provide a focused introduction of data-mining methods that are used in the latter half of the book. These chapters are not intended to be an exhaustive, scholarly treatise on data mining. It is designed only to discuss the methods commonly used in management and engineering design. The real gem of this book lies in Chapters 7–14, where we introduce how to use data-mining methods to solve management and industrial engineering design problems. The details of this book are as follows.

In Chapter 1, we introduce two simple but widely used methods: decision analysis and cluster analysis. Decision analysis is used to make decisions under an un-

certain business environment. Cluster analysis helps us find homogenous objects, called clusters, which are similar and/or well separated.

Chapter 2 interprets the association rules mining method, which is an important topic in data mining. Association rules mining is used to discover association relationships or correlations among a set of objects.

Chapter 3 describes fuzzy modeling and optimization methods. Real-world situations are often not deterministic. There exist various types of uncertainties in social, industrial and economic systems. After introducing basic terminology and various theories on fuzzy sets, this chapter aims to present a brief summary of the theory and methods on fuzzy optimization and tries to give readers a clear and comprehensive understanding of fuzzy modeling and fuzzy optimization.

In Chapter 4, we give an introduction of quadratic programming problems with a type of fuzzy objective and resource constraints. We first introduce a genetic algorithms based interactive approach. Then, an approach is interpreted, which focuses on a symmetric model for a kind of fuzzy nonlinear programming problem by way of a special genetic algorithm with mutation along the weighted gradient direction. Finally, a non-symmetric model for a type of fuzzy nonlinear programming problems with penalty coefficients is described by using a numerical example.

Chapter 5 gives an introduction of basic concepts and algorithms of neural networks and self-organizing maps. The self-organizing maps based method has many practical applications, such as semantic map, diagnosis of speech voicing, solving combinatorial optimization problems, and so on. Several numerical examples are used to show various properties of self-organizing maps.

Chapter 6 introduces an important topic in data mining, privacy-preserving data mining (PPDM), which is one of the newest trends in privacy and security research. It is driven by one of the major policy issues of the information era: the right to privacy. Data are distributed among various parties. Legal and commercial concerns may prevent the parties from directly sharing some sensitive data. How parties collaboratively conduct data mining without breaching data privacy presents a grand challenge. In this chapter, some techniques for privacy-preserving data mining are introduced.

In Chapter 7, decision analysis models are developed to study the benefits from cooperation and leadership in a supply chain. A total of eight cooperation/leadership policies of the leader company are analyzed by using four models. Optimal decisions for the leader company under different cost combinations are analyzed.

Using a decision tree, Chapter 8 characterizes the impact of product global performance on the choice of product architecture during the product development process. We divide product architectures into three categories: modular, hybrid, and integral. This chapter develops analytic models whose objectives are obtaining global performance of a product through a modular/hybrid/integral architecture. Trade-offs between costs and expected benefits from different product architectures are analyzed and compared.

Chapter 9 reviews various cluster analysis methods that have been applied in cellular manufacturing design. We give a comprehensive overview and discussion

for similarity coefficients developed to date for use in solving the cell formation problem. To summarize various similarity coefficients, we develop a classification system to clarify the definition and usage of various similarity coefficients in designing cellular manufacturing systems. Existing similarity (dissimilarity) coefficients developed so far are mapped onto the taxonomy. Additionally, production information-based similarity coefficients are discussed and a historical evolution of these similarity coefficients is outlined. We compare the performance of twenty well-known similarity coefficients. More than two hundred numerical cell formation problems, which are selected from the literature or generated deliberately, are used for the comparative study. Nine performance measures are used for evaluating the goodness of cell formation solutions.

Chapter 10 develops a cluster analysis method to solve a cell formation problem. A similarity coefficient is proposed, which incorporates alternative process routing, operation sequence, operation time, and production volume factors. This similarity coefficient is used to solve a cell formation problem that incorporates various real-life production factors, such as the alternative process routing, operation sequence, operation time, production volume of parts, machine capacity, machine investment cost, machine overload, multiple machines available for machine types and part process routing redesigning cost.

In Chapter 11, we show how to use a fuzzy modeling approach and a genetic-based interactive approach to control a product's quality. We consider a quality function deployment (QFD) design problem that incorporates financial factor and plan uncertainties. A QFD-based integrated product development process model is presented firstly. By introducing some new concepts of planned degree, actual achieved degree, actual primary costs required and actual planned costs, two types of fuzzy nonlinear optimization models are introduced in this chapter. These models not only consider the overall customer satisfaction, but also the enterprise satisfaction with the costs committed to the product.

Chapter 12 introduces a key decision making problem in a supply chain system: inventory control. We establish a new algorithm of inventory classification based on the association rules, in which by using the support-confidence framework the consideration of the cross-selling effect is introduced to generate a new criterion that is then used to rank inventory items. Then, a numerical example is used to explain the new algorithm and empirical experiments are implemented to evaluate its effectiveness and utility, comparing with traditional ABC classification.

In Chapter 13, we describe a technology, surface mountain technology (SMT), which is used in the modern electronics and electronic device industry. A key part for SMT is to construct master data. We propose a method of making master data by using a self-organizing maps learning algorithm and prove such a method is effective not only in judgment accuracy but also in computational feasibility. Empirical experiments are invested for proving the performance of the indicator. Consequently, the continuous weight is effective for the learning evaluation in the process of making the master data.

Chapter 14 describes applications of data mining with privacy-preserving capability, which has been an area gaining researcher attention recently. We introduce applications from various perspectives. Firstly, we present privacy-preserving association rule mining. Then, methods for privacy-preserving classification in data mining are introduced. We also discuss privacy-preserving clustering and a scheme to privacy-preserving collaborative data mining.

Yamagata University, Japan
December 2010

Yong Yin
Ikou Kaku
Jiafu Tang
JianMing Zhu

http://www.springer.com/978-1-84996-337-4

Data Mining
Concepts, Methods and Applications in Management and Engineering Design
Yin, Y.; Kaku, I.; Tang, J.; Zhu, J.
2011, XIV, 312 p., Hardcover
ISBN: 978-1-84996-337-4