教育部高等教育司推荐
国外优秀信息科学与技术系列教学用书

数据挖掘
——概念与技术
（影印版）

DATA MINING
Concepts and Techniques

Jiawei Han
Micheline Kamber

高等教育出版社
Higher Education Press
Morgan Kaufmann Publishers

教育部高等教育司推荐
国外优秀信息科学与技术系列教学用书

数据挖掘

——概念与技术

（影印版）

DATA MINING

Concepts and Techniques

Jiawei Han
Micheline Kamber

高等教育出版社
Morgan Kaufmann Publishers

图字：01-2001-1039 号

Data Mining: Concepts and Techniques

Jiawei Han & Micheline Kamber

Copyright © 2000 by Academic Press
All Rights Reserved
This edition is authorized for sale only in the People's Republic of China (excluding the Special Administrative Regions of Hong Kong and Macau)

图书在版编目(CIP)数据

数据挖掘：概念与技术：英文／（美）韩（Han,J.）
—北京：高等教育出版社，2001
ISBN 7-04-010041-X

Ⅰ.数… Ⅱ.韩… Ⅲ.数据处理－英文 Ⅳ.TP274

中国版本图书馆 CIP 数据核字（2001）第 19784 号

数据挖掘——概念与技术
Jiawei Han 等

出版发行	高等教育出版社		
社　　址	北京市东城区沙滩后街 55 号	**邮政编码**	100009
电　　话	010—64054588	**传　　真**	010—64014048
网　　址	http://www.hep.edu.cn http://www.hep.com.cn		
经　　销	新华书店北京发行所		
印　　刷	北京民族印刷厂		
开　　本	787 × 1092　1/16	**版　　次**	2001 年 5 月第 1 版
印　　张	36	**印　　次**	2001 年10月第 2 次印刷
字　　数	762 000	**定　　价**	35.00 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

版权所有　侵权必究

前　言

20世纪末，以计算机和通信技术为代表的信息科学和技术，对世界的经济、军事、科技、教育、文化、卫生等方面的发展产生了深刻的影响，由此而兴起的信息产业已经成为世界经济发展的支柱。进入21世纪，各国为了加快本国的信息产业，加大了资金投入和政策扶持。

为了加快我国信息产业的进程，在我国《国民经济和社会发展第十个五年计划纲要》中，明确提出“以信息化带动工业化，发挥后发优势，实现社会生产力的跨越式发展。”信息产业的国际竞争将日趋激烈。在我国加入WTO后，我国信息产业将面临国外竞争对手的严峻挑战。竞争成败最终将取决于信息科学和技术人才的多少与优劣。

在20世纪末，我国信息产业虽然得到迅猛发展，但与国际先进国家相比，差距还很大。为了赶上并超过国际先进水平，我国必须加快信息技术人才的培养，特别要培养一大批具有国际竞争能力的高水平的信息技术人才，促进我国信息产业和国家信息化水平的全面提高。为此，教育部高等教育司根据教育部吕福源副部长的意见，在长期重视推动高等学校信息科学和技术的教学的基础上，将实施超前发展战略，采取一些重要举措，加快推动高等学校的信息科学和技术等相关专业的教学工作。在大力宣传、推荐我国专家编著的面向21世纪和“九五”重点的信息科学和技术课程教材的基础上，在有条件的高等学校的某些信息科学和技术课程中推动使用国外优秀教材的影印版进行英语或双语教学，以缩短我国在计算机教学上与国际先进水平的差距，同时也有助于强化我国大学生的英语水平。

为了达到上述目的，在分析一些出版社已影印相关教材，一些学校已试用影印教材进行教学的基础上，教育部高等教育司组织并委托高等教育出版社开展国外优秀信息科学和技术优秀教材及其教学辅助材料的引进研究与影印出版的试点工作。为推动用影印版教材进行教学创造条件。

本次引进的系列教材的影印出版工作，是在对我国高校的信息科学和技术专业的课程与美国高校的进行对比分析的基础上展开的；所影印出版的教材均由我国主要高

校的信息科学和技术专家组成的专家组，从国外近两年出版的大量最新教材中精心筛选评审通过的内容新、有影响的优秀教材；影印教材的定价原则上应与我国大学教材价格相当。

教育部高等教育司将此影印系列教材推荐给高等学校，希望有关教师选用，使用后有什么意见和建议请及时反馈。也希望有条件的出版社，根据影印教材的要求，积极参加此项工作，以便引进更多、更新、更好的外国教材和教学辅助材料。

同时，感谢国外有关出版公司对此项引进工作的配合，欢迎更多的国外公司关心并参与此项工作。

教育部高等教育司

二〇〇一年四月

试读结束：需要全本请在线购买： www.ertongbook.com

To Yandong and Lawrence for your love and encouragement

J.H.

To Erik, Kevan, and Kian for your love and inspiration

M.K.

Foreword

by Jim Gray
Microsoft Research

We are deluged by data—scientific data, medical data, demographic data, financial data, and marketing data. People have no time to look at this data. Human attention has become a precious resource. So, we must find ways to automatically analyze the data, to automatically classify it, to automatically summarize it, to automatically discover and characterize trends in it, and to automatically flag anomalies. This is one of the most active and exciting areas of the database research community. Researchers in areas such as statistics, visualization, artificial intelligence, and machine learning are contributing to this field. The breadth of the field makes it difficult to grasp its extraordinary progress over the last few years.

Jiawei Han and Micheline Kamber have done a wonderful job of organizing and presenting data mining in this very readable textbook. They begin by giving quick introductions to database and data mining concepts with particular emphasis on data analysis. They review the current product offerings by presenting a general framework that covers them all. They then cover in a chapter-by-chapter tour the concepts and techniques that underlie classification, prediction, association, and clustering. These topics are presented with examples, a tour of the best algorithms for each problem class, and pragmatic rules of thumb about when to apply each technique. I found this presentation style to be very readable, and I certainly learned a lot from reading the book. Jiawei Han and Micheline Kamber have been leading contributors to data mining research. This is the text they use with their students to bring them up to speed on the field. The field is evolving very rapidly, but this book is a quick way to learn the basic ideas, and to understand where the field is today. I found it very informative and stimulating, and I expect you will too.

Preface

Our capabilities of both generating and collecting data have been increasing rapidly in the last several decades. Contributing factors include the widespread use of bar codes for most commercial products, the computerization of many business, scientific, and government transactions, and advances in data collection tools ranging from scanned text and image platforms to satellite remote sensing systems. In addition, popular use of the World Wide Web as a global information system has flooded us with a tremendous amount of data and information. This explosive growth in stored data has generated an urgent need for new techniques and automated tools that can intelligently assist us in transforming the vast amounts of data into useful information and knowledge.

This book explores the concepts and techniques of *data mining*, a promising and flourishing frontier in database systems and new database applications. Data mining, also popularly referred to as *knowledge discovery in databases (KDD)*, is the automated or convenient extraction of patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories.

Data mining is a multidisciplinary field, drawing work from areas including database technology, artificial intelligence, machine learning, neural networks, statistics, pattern recognition, knowledge-based systems, knowledge acquisition, information retrieval, high-performance computing, and data visualization. We present the material in this book from a *database perspective*. That is, we focus on issues relating to the feasibility, usefulness, efficiency, and scalability of techniques for the discovery of patterns hidden *in large databases*. As a result, this book is not intended as an introduction to database systems, machine learning, statistics, or other such areas, although we do provide the background necessary in these areas in order to facilitate the reader's comprehension of their respective roles in data mining. Rather, the book is a comprehensive introduction to data mining, presented with database issues in focus. It should be useful for computing science students, application developers, and business professionals, as well as researchers involved in any of the disciplines listed above.

Data mining emerged during the late 1980s, has made great strides during the 1990s, and is expected to continue to flourish into the new millennium. This

book presents an overall picture of the field from a database researcher's point of view, introducing interesting data mining techniques and systems, and discussing applications and research directions. An important motivation for writing this book was the need to build an organized framework for the study of data mining—a challenging task owing to the extensive multidisciplinary nature of this fast developing field. We hope that this book will encourage people with different backgrounds and experiences to exchange their views regarding data mining so as to contribute toward the further promotion and shaping of this exciting and dynamic field.

To the Teacher

This book is designed to give a broad, yet in-depth overview of the field of data mining. You will find it useful for teaching a course on data mining at an advanced undergraduate level or the first-year graduate level. In addition, individual chapters may be included as material for courses on selected topics in database systems or in artificial intelligence. We have tried to make the chapters as self-contained as possible so that you are not confined to reading each chapter in sequence. For a course taught at the undergraduate level, you might use Chapters 1 through 8 as the core course material. Remaining class material may be selected from among the more advanced topics described in Chapters 9 and 10. For a graduate-level course, you may choose to cover the entire book in one semester.

Each chapter ends with a set of exercises, suitable as assigned homework. The exercises are either short questions that test basic mastery of the material covered, or longer questions that require analytical thinking.

To the Student

We hope that this textbook will spark your interest in the fresh, yet evolving field of data mining. We have attempted to present the material in a clear manner, with careful explanation of the topics covered. Each chapter ends with a summary describing the main points. We have included many figures and illustrations throughout the text in order to make the book more enjoyable and "reader-friendly." Although this book was designed as a textbook, we have tried to organize it so that it will also be useful to you as a reference book or handbook, should you later decide to pursue a career in data mining.

What do you need to know in order to read this book?

- You should have some knowledge of the concepts and terminology associated with database systems. However, we do try to provide enough background of the basics in database technology, so that if your memory is a bit rusty, you will

not have trouble following the discussions in the book. You should have some knowledge of database querying, although knowledge of any specific query language is not required.

- You should have some programming experience. In particular, you should be able to read pseudocode, and understand simple data structures such as multidimensional arrays.
- It will be helpful to have some preliminary background in statistics, machine learning, or pattern recognition. However, we will familiarize you with the basic concepts of these areas that are relevant to data mining from a database perspective.

To the Professional

This book was designed to cover a broad range of topics in the field of data mining. As a result, it is an excellent handbook on the subject. Because each chapter is designed to be as stand-alone as possible, you can focus on the topics that most interest you. Much of the book is suited to applications programmers or information service managers like yourself who wish to learn about the key ideas of data mining on their own.

The techniques and algorithms presented are of practical utility. Rather than selecting algorithms that perform well on small "toy" databases, the algorithms described in the book are geared for the discovery of data patterns hidden in large, real databases. In Chapter 10, we briefly discuss data mining systems in commercial use, as well as promising research prototypes. Each algorithm presented in the book is illustrated in pseudocode. The pseudocode is similar to the C programming language, yet is designed so that it should be easy to follow by programmers unfamiliar with C or C++. If you wish to implement any of the algorithms, you should find the translation of our pseudocode into the programming language of your choice to be a fairly straightforward task.

Organization of the Book

The book is organized as follows.

Chapter 1 provides an introduction to the multidisciplinary field of data mining. It discusses the evolutionary path of database technology that has led to the need for data mining, and the importance of its application potential. The basic architecture of data mining systems is described, and a brief introduction to the concepts of database systems and data warehouses is given. A detailed classification of data mining tasks is presented, based on the different kinds of knowledge

to be mined. A classification of data mining systems is presented, and major challenges in the field are discussed.

Chapter 2 is an introduction to data warehouses and OLAP (On-Line Analytical Processing). Topics include the concept of data warehouses and multidimensional databases, the construction of data cubes, the implementation of on-line analytical processing, and the relationship between data warehousing and data mining.

Chapter 3 describes techniques for preprocessing the data prior to mining. Methods of data cleaning, data integration and transformation, and data reduction are discussed, including the use of concept hierarchies for dynamic and static discretization. The automatic generation of concept hierarchies is also described.

Chapter 4 introduces the primitives of data mining that define the specification of a data mining task. It describes a data mining query language (DMQL) and provides examples of data mining queries. Other languages are also described, as well as the construction of graphical user interfaces and data mining architectures.

Chapter 5 describes techniques for concept description, including characterization and discrimination. An attribute-oriented generalization technique is introduced, as well as its different implementations including a generalized relation technique and a multidimensional data cube technique. Several forms of knowledge presentation and visualization are illustrated. Relevance analysis is discussed. Methods for class comparison at multiple abstraction levels and methods for the extraction of characteristic rules and discriminant rules with interestingness measurements are presented. In addition, statistical measures for descriptive mining are discussed.

Chapter 6 presents methods for mining association rules in transaction databases as well as relational databases and data warehouses. It includes a classification of association rules, a presentation of the basic Apriori algorithm and its variations, and techniques for mining multilevel association rules, multidimensional association rules, quantitative association rules, and correlation rules. A new technique called frequent pattern growth is introduced, which mines frequent patterns without candidate set generation. Strategies for finding interesting rules by constraint-based mining and the use of interestingness measures to focus the rule search are also described.

Chapter 7 describes methods for data classification and prediction, including decision tree induction, Bayesian classification, the neural network technique of backpropagation, k-nearest neighbor classifiers, case-based reasoning, genetic algorithms, rough set theory, and fuzzy set approaches. Classification based on concepts from association rule mining is presented. Methods of regression are introduced, and issues regarding classifier accuracy are discussed.

Chapter 8 describes methods of cluster analysis. It first introduces the concept of data clustering and then presents several major data clustering approaches, including partition-based clustering, hierarchical clustering, and model-based clustering. Methods for clustering continuous data, discrete data, and data in mul-

试读结束：需要全本请在线购买： www.ertongbook.com

tidimensional data cubes are presented. The scalability of clustering algorithms is discussed in detail.

Chapter 9 discusses methods for data mining in advanced database systems. It includes data mining in object-oriented databases, spatial databases, multimedia databases, time-series databases, text databases, and the World Wide Web.

Finally, in Chapter 10, we summarize the concepts presented in this book and discuss applications of data mining and some challenging research issues.

Throughtout the text, italic is used to emphasize terms that are defined, while bold is used to highlight main ideas.

Errors

It is likely that this book may contain typos, errors, or omissions. If you notice any errors, have suggestions regarding additional exercises, or have other constructive criticism, we would be very happy to hear from you. We welcome and appreciate your suggestions. You can send your comments to

Data Mining: Concepts and Techniques
Intelligent Database Systems Research Laboratory
School of Computing Science
Simon Fraser University
Burnaby, British Columbia
Canada V5A 1S6
Fax: (604) 291-3045

Alternatively, you can use electronic mail to submit bug reports, request a list of known errors, or make constructive suggestions. To receive instructions, send e-mail to *dmbook@cs.sfu.ca* with "Subject: help" in the message header. We regret that we cannot personally respond to all e-mail messages. The errata of the book and other updated information related to the book can be found by referencing the Web address *www.cs.sfu.ca/~han/DM_Book*.

Acknowledgments

We would like to express our sincere thanks to all those who have worked or are currently working with us on data mining related research and/or the DBMiner project, or have provided us with various support in data mining. These include Rakesh Agrawal, Stella Atkins, Yvan Bedard, Binay Bhattacharya, Dora (Yandong) Cai, Nick Cercone, Surajit Chaudhuri, Sonny H. S. Chee, Jianping Chen, Ming-Syan Chen, Qing Chen, Qiming Chen, Shan Cheng, David Cheung, Shi Cong, Son Dao, Umeshwar Dayal, James Delgrande, Guozhu Dong, Carole Edwards,

Max Egenhofer, Martin Ester, Usama Fayyad, Ling Feng, Ada Fu, Yongjian Fu, Daphne Gelbart, Randy Goebel, Jim Gray, Robert Grossman, Wan Gong, Yike Guo, Eli Hagen, Howard Hamilton, Jing He, Larry Henschen, Jean Hou, Mei-Chun Hsu, Kan Hu, Haiming Huang, Yue Huang, Julia Itskevitch, Wen Jin, Tiko Kameda, Hiroyuki Kawano, Rizwan Kheraj, Eddie Kim, Won Kim, Krzysztof Koperski, Hans-Peter Kriegel, Vipin Kumar, Laks V.S. Lakshmanan, Joyce Man Lam, James Lau, Deyi Li, George (Wenmin) Li, Jin Li, Ze-Nian Li, Nancy Liao, Gang Liu, Junqiang Liu, Ling Liu, Alan (Yijun) Lu, Hongjun Lu, Tong Lu, Wei Lu, Xuebin Lu, Wo-Shun Luk, Heikki Mannila, Runying Mao, Abhay Mehta, Gabor Melli, Alberto Mendelzon, Tim Merrett, Harvey Miller, Drew Miners, Behzad Mortazavi-Asl, Richard Muntz, Raymond T. Ng, Vicent Ng, Shojiro Nishio, Beng-Chin Ooi, Tamer Ozsu, Jian Pei, Gregory Piatetsky-Shapiro, Helen Pinto, Fred Popowich, Amynmohamed Rajan, Peter Scheuermann, Shashi Shekhar, Wei-Min Shen, Avi Silberschatz, Evangelos Simoudis, Nebojsa Stefanovic, Yin Jenny Tam, Simon Tang, Zhaohui Tang, Dick Tsur, Anthony K. H. Tung, Ke Wang, Wei Wang, Zhaoxia Wang, Tony Wind, Lara Winstone, Ju Wu, Betty (Bin) Xia, Cindy M. Xin, Xiaowei Xu, Qiang Yang, Yiwen Yin, Clement Yu, Jeffrey Yu, Philip S. Yu, Osmar R. Zaiane, Carlo Zaniolo, Shuhua Zhang, Zhong Zhang, Yvonne Zheng, Xiaofang Zhou, and Hua Zhu. We are also grateful to Jean Hou, Helen Pinto, Lara Winstone, and Hua Zhu for their help with some of the original figures in this book, and to Eugene Belchev for his careful proofreading of each chapter.

We also wish to thank Diane Cerra, our Executive Editor at Morgan Kaufmann Publishers, for her enthusiasm, patience, and support during our writing of this book, as well as Howard Severson, our Production Editor, and his staff for their conscientious efforts regarding production. We are indebted to all of the reviewers for their invaluable feedback. Finally, we thank our families for their wholehearted support throughout this project.

Contents

Foreword vii

Preface xix

Chapter 1 **Introduction 1**

1.1 **What Motivated Data Mining? Why Is It Important? 1**

1.2 **So, What Is Data Mining? 5**

1.3 **Data Mining—On What Kind of Data? 10**

1.3.1 Relational Databases 10

1.3.2 Data Warehouses 12

1.3.3 Transactional Databases 15

1.3.4 Advanced Database Systems and Advanced Database Applications 16

1.4 **Data Mining Functionalities—What Kinds of Patterns Can Be Mined? 21**

1.4.1 Concept/Class Description: Characterization and Discrimination 21

1.4.2 Association Analysis 23

1.4.3 Classification and Prediction 24

1.4.4 Cluster Analysis 25

1.4.5 Outlier Analysis 25

1.4.6 Evolution Analysis 26

1.5 **Are All of the Patterns Interesting? 27**

1.6 **Classification of Data Mining Systems 28**

1.7 **Major Issues in Data Mining 30**

1.8 **Summary 33**

Exercises 34

Bibliographic Notes 35

Chapter 2 **Data Warehouse and OLAP Technology for Data Mining 39**

2.1 **What Is a Data Warehouse? 39**

2.1.1 Differences between Operational Database Systems and Data Warehouses 42

2.1.2 But, Why Have a Separate Data Warehouse? 44

2.2 **A Multidimensional Data Model 44**
2.2.1 From Tables and Spreadsheets to Data Cubes 45
2.2.2 Stars, Snowflakes, and Fact Constellations: Schemas for Multidimensional Databases 48
2.2.3 Examples for Defining Star, Snowflake, and Fact Constellation Schemas 52
2.2.4 Measures: Their Categorization and Computation 54
2.2.5 Introducing Concept Hierarchies 56
2.2.6 OLAP Operations in the Multidimensional Data Model 58
2.2.7 A Starnet Query Model for Querying Multidimensional Databases 61
2.3 **Data Warehouse Architecture 62**
2.3.1 Steps for the Design and Construction of Data Warehouses 63
2.3.2 A Three-Tier Data Warehouse Architecture 65
2.3.3 Types of OLAP Servers: ROLAP versus MOLAP versus HOLAP 69
2.4 **Data Warehouse Implementation 71**
2.4.1 Efficient Computation of Data Cubes 71
2.4.2 Indexing OLAP Data 79
2.4.3 Efficient Processing of OLAP Queries 81
2.4.4 Metadata Repository 83
2.4.5 Data Warehouse Back-End Tools and Utilities 84
2.5 **Further Development of Data Cube Technology 85**
2.5.1 Discovery-Driven Exploration of Data Cubes 85
2.5.2 Complex Aggregation at Multiple Granularities: Multifeature Cubes 89
2.5.3 Other Developments 92
2.6 **From Data Warehousing to Data Mining 93**
2.6.1 Data Warehouse Usage 93
2.6.2 From On-Line Analytical Processing to On-Line Analytical Mining 95
2.7 **Summary 98**
Exercises 99
Bibliographic Notes 103

Chapter 3 **Data Preprocessing 105**
3.1 **Why Preprocess the Data? 105**
3.2 **Data Cleaning 109**
3.2.1 Missing Values 109
3.2.2 Noisy Data 110
3.2.3 Inconsistent Data 112
3.3 **Data Integration and Transformation 112**
3.3.1 Data Integration 112
3.3.2 Data Transformation 114

3.4 **Data Reduction 116**
3.4.1 Data Cube Aggregation 117
3.4.2 Dimensionality Reduction 119
3.4.3 Data Compression 121
3.4.4 Numerosity Reduction 124
3.5 **Discretization and Concept Hierarchy Generation 130**
3.5.1 Discretization and Concept Hierarchy Generation for Numeric Data 132
3.5.2 Concept Hierarchy Generation for Categorical Data 138
3.6 **Summary 140**
Exercises 141
Bibliographic Notes 142

Chapter 4 **Data Mining Primitives, Languages, and System Architectures 145**
4.1 **Data Mining Primitives: What Defines a Data Mining Task? 146**
4.1.1 Task-Relevant Data 148
4.1.2 The Kind of Knowledge to be Mined 150
4.1.3 Background Knowledge: Concept Hierarchies 151
4.1.4 Interestingness Measures 155
4.1.5 Presentation and Visualization of Discovered Patterns 157
4.2 **A Data Mining Query Language 159**
4.2.1 Syntax for Task-Relevant Data Specification 160
4.2.2 Syntax for Specifying the Kind of Knowledge to be Mined 162
4.2.3 Syntax for Concept Hierarchy Specification 165
4.2.4 Syntax for Interestingness Measure Specification 166
4.2.5 Syntax for Pattern Presentation and Visualization Specification 167
4.2.6 Putting It All Together—An Example of a DMQL Query 167
4.2.7 Other Data Mining Languages and the Standardization of Data Mining Primitives 169
4.3 **Designing Graphical User Interfaces Based on a Data Mining Query Language 170**
4.4 **Architectures of Data Mining Systems 171**
4.5 **Summary 174**
Exercises 174
Bibliographic Notes 176

Chapter 5 **Concept Description: Characterization and Comparison 179**
5.1 **What Is Concept Description? 179**
5.2 **Data Generalization and Summarization-Based Characterization 181**

5.2.1 Attribute-Oriented Induction 182
5.2.2 Efficient Implementation of Attribute-Oriented Induction 187
5.2.3 Presentation of the Derived Generalization 190

5.3 **Analytical Characterization: Analysis of Attribute Relevance 194**
5.3.1 Why Perform Attribute Relevance Analysis? 195
5.3.2 Methods of Attribute Relevance Analysis 196
5.3.3 Analytical Characterization: An Example 198

5.4 **Mining Class Comparisons: Discriminating between Different Classes 200**
5.4.1 Class Comparison Methods and Implementations 201
5.4.2 Presentation of Class Comparison Descriptions 204
5.4.3 Class Description: Presentation of Both Characterization and Comparison 206

5.5 **Mining Descriptive Statistical Measures in Large Databases 208**
5.5.1 Measuring the Central Tendency 209
5.5.2 Measuring the Dispersion of Data 210
5.5.3 Graph Displays of Basic Statistical Class Descriptions 213

5.6 **Discussion 217**
5.6.1 Concept Description: A Comparison with Typical Machine Learning Methods 218
5.6.2 Incremental and Parallel Mining of Concept Description 220

5.7 **Summary 220**
Exercises 222
Bibliographic Notes 223

Chapter 6 **Mining Association Rules in Large Databases 225**

6.1 **Association Rule Mining 226**
6.1.1 Market Basket Analysis: A Motivating Example for Association Rule Mining 226
6.1.2 Basic Concepts 227
6.1.3 Association Rule Mining: A Road Map 229

6.2 **Mining Single-Dimensional Boolean Association Rules from Transactional Databases 230**
6.2.1 The Apriori Algorithm: Finding Frequent Itemsets Using Candidate Generation 230
6.2.2 Generating Association Rules from Frequent Itemsets 236
6.2.3 Improving the Efficiency of Apriori 236
6.2.4 Mining Frequent Itemsets without Candidate Generation 239
6.2.5 Iceberg Queries 243

6.3 **Mining Multilevel Association Rules from Transaction Databases 244**

试读结束：需要全本请在线购买： www.ertongbook.com