



BOOK REVIEW

Data-mining and Uncertain Reasoning: An Integrated Approach

By Z Chen, published by John Wiley, New York, 2001 Hardback, ISBN 0471388785

Reviewed by: Xiaohua Hu

College of Information Science and Technology,
Drexel University, Philadelphia, PA 19104, USA

Information Visualization (2003) 2, 78–79.
doi:10.1057/palgrave.ivs.9500041

Data-mining is a process of extracting interesting, useful and actionable patterns from large data sets. It has become a very important field and has attracted a lot of attention from researchers and practitioners with different training and education background since its inception. Uncertain reasoning approaches such as fuzzy set theory, rough set theory, neural network and genetic algorithms are powerful computational tools for data analysis and data-mining due to its greater expressiveness to deal with uncertainty information in the data sets than most rule induction algorithms and have been playing an increasingly important role in the data-mining procedure. Nowadays, there are many good books on uncertain reasoning and data-mining separately, the depth and coverage of these books are impressive, but they do not provide an integrated view of the complicated and broad data-mining and uncertain reasoning fields. This book fills the gap in this regard. This book is so far the only one that integrates two closely related research areas: data-mining and uncertain reasoning. It provides a holistic view of the data-mining field from the uncertain reasoning perspective.

The coverage of this book is impressive. The bulk of the text is devoted to detailed explanations of the various types of data-mining algorithms, including decision tree, association rules, Bayesian network, artificial neural network all the way to fuzzy set, rough set and genetic algorithms. Other sections of the book serve to place data-mining in perspective relative to other techniques for exploiting data. *'Data-mining and Uncertain Reasoning: An Integrated Approach'* reviews the common ground between the two disciplines. In this book, the author takes an integrated approach to combine two important topics in a seamless way. The book covers a wide range of important topics in the data-mining fields: data-mining objective, life cycle, tasks, infrastructure and application, uncertain reasoning (fuzzy set, rough set, neural network and genetic algorithms), the relation between them, case study as well as future directions. The book is organized in three parts. Part 1 consists of Chapters 1–4. Part 2 comprises Chapters 5–7 and Part 3 Chapters 8 and 9. In addition, the materials in each chapter are organized into two categories: basic and advanced, to meet the needs of different readers with different levels of technical skills and training background. Each chapter starts with a general overview to serve as a guideline for the entire chapter. Business readers who are interested in general idea of data-mining may focus on the basic part, while technical-oriented readers may find the advanced part very enjoyable.

The chapters in Part 1 take a step-by-step approach to developing an understanding of the methodologies of data-mining and uncertain reasoning. The chapters progress from a discussion of historical back-



grounds and important concepts, and related enabling techniques to the relation between data-mining and uncertain reasoning. Chapter 1 presents the organization of the entire book, why this book is needed and how to use this book. Chapter 2 begins with basic concepts of data-mining, definition of data-mining, the difference of data-mining from statistical data analysis, the data-mining primitives and process, important issues of data-mining and overview of knowledge discovery task. It then discusses the scalability issues of data-mining methods and theoretical foundation of data-mining and ends with research issues, trends and directions of data-mining. Since most of the data-mining systems rely on the data in the data warehouse as the source for data-mining and the data analyst often uses on-line analytical processing (OLAP) tools to explore the basic relations in the data set before mining the data set, an introduction of these two topics is very useful and can help readers gain a better understanding of the data-mining process. Chapter 3 is devoted to data warehouse, OLAP, role of meta data, integration of OLAP with data-mining, parallel data-mining integrated with data warehousing. Chapter 4 studies the uncertainty issues associated with data manipulation, source and types of uncertainty, uncertainty in data and schema integration, existing work in dealing with data and knowledge uncertainty and various approaches to uncertain reasoning in the data-mining procedure.

Part 2 (Chapters 5–7) provides detailed examinations of various data-mining techniques and methods. The discussions focus on the technical details of various algorithms. Chapter 5 presents the important data-mining tasks to knowledge discovery. This chapter begins with discussion of AI, symbol-based machine learning, induction and generalization, and introduces the attribute-oriented generation method. It then examines in great detail the three most important data-mining tasks: classification, clustering and association rules. Chapter 6 is devoted to two important and popular uncertain reasoning approaches and their applications in the data-mining procedures: Bayesian networks and artificial neural networks (ANNs). This chapter discusses the fundamentals of both approaches and then extends to how to apply them in data-mining procedures. Chapter 7 discusses the important aspects of fuzzy set, rough set and genetic algorithms. The discussion of these approaches focuses on the differences among them, and how each of them is used in the data-mining procedure.

Perhaps the very most practical chapters are in Part 3. Chapter 8 is, in my opinion, the most interesting chapter of the book. This chapter presents several case studies in various applications: from bank credit applications, incorporation of OLAP with data-mining, multiple level association rule mining, *k*-means algorithms with initial points refining to data-mining reduction approach. Chapter 9 studies data-mining techniques to intelligent concept query resolution with uncertainty, this topic is a very promising direction for data-mining research and application.

The purpose of this book is to introduce useful techniques for data-mining and uncertain reasoning, present a holistic perspective of uncertain reasoning for data-mining. *Data-mining and Uncertain Reasoning: An Integrated Approach* can be used as a nice reference book for both beginners and experts in the data-mining field. It is a good book for business, technical professional, students and data-mining researchers. For the business and technical professionals, this book explains the power of data-mining and how data-mining can help solve some challenging business problems. For students and data-mining researchers, it provides with a comprehensive introduction to the background, methodologies and applications of data-mining with uncertain reasoning. The book is written in a semiformal, 'how-to' format that is easy and enjoyable reading. The case studies provide examples of 'real world' types of challenging problems that data-mining strives to solve.

However, the book can be improved in the following aspects to enhance the readability. The review chapters are relatively short, considering the scope and complexity of the research topics addressed by the book. Therefore, for a good understanding of the book, the reader may need to have some background in data-mining, fuzzy sets, rough sets, neural network and statistics. If the reader is not familiar with any of these areas, he should probably read another book about those topics before starting to read this book. Another limitation is that this book has no exercise for students, so maybe it is not suitable for use as a textbook for undergraduate level data-mining course.

In short, *Data-mining and Uncertain Reasoning: An Integrated Approach* is quite readable, even for those without a database, AI or statistics background, and is definitely recommended for anyone wanting to embark on more than a casual study in this exciting and fast growing field.