Lecture Notes in Artificial Intelligence 2307

Subseries of Lecture Notes in Computer Science
Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science
Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Tokyo

Chengqi Zhang Shichao Zhang

Association Rule Mining

Models and Algorithms

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Authors

Chengqi Zhang
Shichao Zhang
University of Technology, Sydney, Faculty of Information Technology
P.O. Box 123 Broadway, Sydney, NSW 2007 Australia
E-mail: {chengqi,zhangsc}@it.uts.edu.au

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Zhang, Chengqi:
Association rule mining : models and algorithms / Chengqi Zhang ;
Shichao Zhang. - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ;
London ; Milan ; Paris ; Tokyo : Springer, 2002
(Lecture notes in computer science ; Vol. 2307 : Lecture notes in
artificial intelligence)
ISBN 3-540-43533-6

CR Subject Classification (1998): I.2.6, I.2, H.2.8, H.2, H.3, F.2.2

ISSN 0302-9743
ISBN 3-540-43533-6 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Typesetting: Camera-ready by author, data conversion by Boller Mediendesign
Printed on acid-free paper SPIN: 10846539 06/3142 5 4 3 2 1 0

Preface

Association rule mining is receiving increasing attention. Its appeal is due, not only to the popularity of its parent topic 'knowledge discovery in databases and data mining', but also to its neat representation and understandability. The development of association rule mining has been encouraged by active discussion among communities of users and researchers. All have contributed to the formation of the technique with a fertile exchange of ideas at important forums or conferences, including SIGMOD, SIGKDD, AAAI, IJCAI, and VLDB. Thus association rule mining has advanced into a mature stage, supporting diverse applications such as data analysis and predictive decisions.

There has been considerable progress made recently on mining in such areas as quantitative association rules, causal rules, exceptional rules, negative association rules, association rules in multi-databases, and association rules in small databases. These continue to be future topics of interest concerning association rule mining. Though the association rule constitutes an important pattern within databases, to date there has been no specilized monograph produced in this area. Hence this book focuses on these interesting topics.

The book is intended for researchers and students in data mining, data analysis, machine learning, knowledge discovery in databases, and anyone else who is interested in association rule mining. It is also appropriate for use as a text supplement for broader courses that might also involve knowledge discovery in databases and data mining.

The book consists of eight chapters, with bibliographies after each chapter. Chapters 1 and 2 lay a common foundation for subsequent material. This includes the preliminaries on data mining and identifying association rules, as well as necessary concepts, previous efforts, and applications. The later chapters are essentially self-contained and may be read selectively, and in any order. Chapters 3, 4, and 5 develop techniques for discovering hidden patterns, including negative association rules and causal rules. Chapter 6 presents techniques for mining very large databases, based on instance selection. Chapter 7 develops a new technique for mining association rules in databases which utilizes external knowledge, and Chapter 8 presents a summary of the previous chapters and demonstrates some open problems.

Beginners should read Chapters 1 and 2 before selectively reading other chapters. Although the open problems are very important, techniques in other chapters may be helpful for experienced readers who want to attack these problems.

January 2002 *Chengqi Zhang and Shichao Zhang*

Acknowledgments

We are deeply indebted to many colleagues for the advice and support they gave during the writing of this book. We are especially grateful to Alfred Hofmann for his efforts in publishing this book with Springer-Verlag. And we thank the anonymous reviewers for their detailed constructive comments on the proposal of this work.

For many suggested improvements and discussions on the material, we thank Professor Geoffrey Webb, Mr. Zili Zhang, and Ms. Li Liu from Deakin University; Professor Huan Liu from Arizona State University, Professor Xindong Wu from Vermont University, Professor Bengchin Ooi and Dr. Kianlee Tan from the National University of Singapore, Dr. Hong Liang and Mr. Xiaowei Yan from Guangxi Normal University, Professor Xiaopei Luo from the Chinese Academy of Sciences, and Professor Guoxi Fan from the Education Bureau of Quanzhou.

Contents