Frequent Pattern Mining

Charu C. Aggarwal • Jiawei Han
Editors

Frequent Pattern Mining

Springer

Editors
Charu C. Aggarwal
IBM T. J. Watson Research Center
Yorktown Heights
New York
USA

Jiawei Han
University of Illinois at Urbana-Champaign
Urbana
Illinois
USA

ISBN 978-3-319-07820-5 ISBN 978-3-319-07821-2 (eBook)
DOI 10.1007/978-3-319-07821-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014944536

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The field of data mining has four main "super-problems" corresponding to clustering, classification, outlier analysis, and frequent pattern mining. Compared to the other three problems, the frequent pattern mining model for formulated relatively recently. In spite of its shorter history, frequent pattern mining is considered the marquee problem of data mining. The reason for this is that interest in the data mining field increased rapidly soon after the seminal paper on association rule mining by Agrawal, Imielinski, and Swami. The earlier data mining conferences were often dominated by a large number of frequent pattern mining papers. This is one of the reasons that frequent pattern mining has a very special place in the data mining community. At this point, the field of frequent pattern mining is considered a mature one.

While the field has reached a relative level of maturity, very few books cover different aspects of frequent pattern mining. Most of the existing books are either too generic or do not cover frequent pattern mining in an exhaustive way. A need exists for an exhaustive book on the topic that can cover the different nuances in an exhaustive way.

This book provides comprehensive surveys in the field of frequent pattern mining. Each chapter is designed as a survey that covers the key aspects of the field of frequent pattern mining. The chapters are typically of the following types:

- *Algorithms*: In these cases, the key algorithms for frequent pattern mining are explored. These include join-based methods such as *Apriori*, and pattern-growth methods.
- *Variations*: Many variations of frequent pattern mining such as interesting patterns, negative patterns, constrained pattern mining, or compressed patterns are explored in these chapters.
- *Scalability*: The large sizes of data in recent years has led to the need for big data and streaming frameworks for frequent pattern mining. Frequent pattern mining algorithms need to be modified to work with these advanced scenarios.
- *Data Types*: Different data types lead to different challenges for frequent pattern mining algorithms. Frequent pattern mining algorithms need to be able to work with complex data types, such as temporal or graph data.

- *Applications*: In these chapters, different applications of frequent pattern mining are explored. These includes the application of frequent pattern mining methods to problems such as clustering and classification. Other more complex algorithms are also explored.

This book is, therefore, intended to provide an overview of the field of frequent pattern mining, as it currently stands. It is hoped that the book will serve as a useful guide for students, researchers, and practitioners.

Contents

Contributors

Charu C. Aggarwal IBM T. J. Watson Research Center, Yorktown Heights, NY, USA

Gagan Agrawal Ohio State University, Columbus, OH, USA

Luiza Antonie University of Guelph, Guelph, Canada

David C. Anastasiu Department of Computer Science and Engineering, University of Minnesota, Minneapolis, USA

Ira Assent Department of Computer Science, Aarhus University, Aarhus, Denmark

Mansurul A. Bhuiyan Indiana University–Purdue University, Indianapolis, IN, USA

Hong Cheng Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong, China

Aris Gkoulalas-Divanis IBM Research-Ireland, Damastown Industrial Estate, Mulhuddart, Dublin, Ireland

Jiawei Han University of Illinois at Urbana-Champaign, Urbana, IL, USA

Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, USA

Jayant Haritsa Database Systems Lab, Indian Institute of Science (IISc), Bangalore, India

Mohammad Al Hasan Indiana University–Purdue University, Indianapolis, IN, USA

Jeremy Iverson Department of Computer Science and Engineering, University of Minnesota, Minneapolis, USA

Ruoming Jin Kent State University, Kent, OH, USA

Murat Kantarcioglu UTD Data Security and Privacy Lab, University of Texas at Dallas, Texas, USA

Victor E. Lee John Carroll University, University Heights, OH, USA

Matthijs van Leeuwen KU Leuven, Leuven, Belgium

Carson Kai-Sang Leung University of Manitoba,Winnipeg, MB, Canada

Jundong Li University of Alberta, Alberta, Canada

Zhenhui Li Pennsylvania State University, University Park, USA

George Karypis Department of Computer Science and Engineering, University of Minnesota, Minneapolis, USA

Siegfried Nijssen KU Leuven, Leuven, Belgium

Universiteit Leiden, Leiden, The Netherlands

Jian Pei Simon Fraser University, Burnaby, BC, Canada

Shaden Smith Department of Computer Science and Engineering, University of Minnesota, Minneapolis, USA

Wei Shen Tsinghua University, Beijing, China

Nikolaj Tatti HIIT, Department of Information and Computer Science, Aalto University, Helsinki, Finland

Jilles Vreeken Max-Planck Institute for Informatics and Saarland University, Saarbrücken, Germany

Jianyong Wang Tsinghua University, Beijing, China

Xifeng Yan Department of Computer Science, University of California at Santa Barbara, Santa Barbara, USA

Osmar Zaiane University of Alberta, Alberta, Canada

Feida Zhu Singapore Management University, Singapore, Singapore

Albrecht Zimmermann INSA Lyon, Villeurbanne CEDEX, France

Arthur Zimek Ludwig-Maximilians-Universität München, Munich, Germany