Preface

SUMMARY

This book examines the post-analysis and post-mining of association rules to find useful knowledge from a large number of discovered rules and presents a systematic view of the above topic. It introduces up-to-date research on extracting useful knowledge from a large number of discovered association rules, and covers interestingness, post-mining, rule selection, summarization, representation and visualization of association rules, as well as new forms of association rules and new trends of association rule mining.

BACKGROUND

As one of the key techniques for data mining, association rule mining was first proposed in 1993, and is today widely used in many applications. An association rule is designed in the form of A→B, where A and B are items or itemsets, for example, beer→diaper. There are often a huge number of association rules discovered from a dataset, and it is sometimes very difficult for a user to identify interesting and useful ones. Therefore, it is important to remove insignificant rules, prune redundancy, summarize, post-mine, and visualize the discovered rules. Moreover, the discovered association rules are in the simple form of A→B, from which the information we can get is very limited. Some recent research has focused on new forms of association rules, such as combined association rules, class association rules, quantitative association rules, contrast patterns and multi-dimensional association rules.

Although there have already been a quite few publications on the post-analysis and post-mining of association rules, there are no books specifically on the above topic. Therefore, we have edited this book to provide a collection of work on the post-mining of association rules and present a whole picture of the post-mining stage of association rule mining.

OBJECTIVES AND SIGNIFICANCE

The objectives of this book are to emphasize the importance of post-mining of association rules, to show a whole picture on the post-mining of association rules, and to present the up-to-date progress of the research on how to extract useful knowledge from a large number of discovered association rules.

The unique characteristic of this book is the comprehensive collection of the current research on post-mining and summarization of association rules and new trends of association rules. It aims to answer the question "we have discovered many association rules, and so what?" It presents readers what we can do or shall do to extract useful and actionable knowledge after discovering a large number of association

rules, instead of algorithms or models for mining association rules themselves. It presents academia with a whole picture of the current research progress on post-mining and summarization of association rules. It may help industry to learn from the ideas and apply them to find useful and actionable knowledge in real-world applications. This book also aims to expand the research on association rules to new areas, such as new forms of association rules. The ideas of post-analysis may also be used in the step of association rule mining and help to make new efficient algorithms for mining more useful association rules.

TARGET AUDIENCES

This book is aimed at researchers, postgraduate students, and practitioners in the field of data mining. For researchers whose interests include data mining, this book presents them with a survey of techniques for post-mining of association rules, the up-to-date research progress and the emerging trends/directions in this area. It may spark new ideas on applying other techniques in data mining, machine learning, statistics, and so forth, to the post-mining phase of association rules, or using the post-mining techniques for association rules to tackle the problems in other fields.

For postgraduate students who are interested in data mining, this book presents an overview of association rule techniques and introduces the origin, interestingness, redundancy, visualization, and maintenance of association rules, as well as associative classification and new forms of association rules. It presents not only the post-mining stage of association rules, but also many techniques that are actually used in association rule mining procedure.

For data miners from the industry, this book provides techniques and methodologies for extracting useful and interesting knowledge from a huge number of association rules learned in a data mining practice. It presents a whole picture of what to do after association rule mining and advanced techniques to post-mine the learned rules. Moreover, it also presents a number of real-life case studies and applications, which may help data miners to design and develop their own data mining projects.

However, the audiences are not limited to those interested in association rules, because the post-mining of association rules involves visualization, clustering, classification and many other techniques of data mining, statistics and machine learning, which are actually beyond association rule mining itself.

ORGANIZATION

This book is composed of six sections. **Section I** gives an introduction to association rules and the current research in the related topics, including the preliminary of association rules and the classic algorithms for association rule mining. **Section II** presents three techniques on using interestingness measures to select useful association rules. **Section III** presents four techniques for the post-processing of associations. **Section IV** presents two techniques for selecting high-quality rules for associative classification. **Section V** discusses three techniques for visualization and representation of association rules. **Section VI** presents the maintenance of association rules and new forms of rules.

Section I presents an introduction to association rule techniques. In **Chapter I**, Paul D. McNicholas and Yanchang Zhao discuss the origin of association rules and the functions by which association rules are traditionally characterised. The formal definition of an association rule, and its support, confidence and lift are presented, and the techniques for rule generation are introduced. It also discusses negations and negative association rules, rule pruning, the measures of interestingness, and the post-mining stage of the association rule paradigm.

Section II studies how to identify interesting rules. In **Chapter II**, Mirko Boettcher, Georg Ruß, Detlef Nauch, and Rudolf Kruse presented a unified view on assessing rule interestingness with the combination of rule change mining and relevance feedback. Rule change mining extends standard association rule mining by generating potentially interesting time-dependent features for an association rule during post-mining, and the existing textual description of a rule and those newly derived objective features are combined by using relevance feedback methods from information retrieval. The proposed technique yields a powerful, intuitive way for exploring the typically vast set of association rules.

Chapter III by Solange Oliveira Rezende et al., presents a new methodology for combining data-driven and user-driven evaluation measures to identify interesting rules. Both data-driven (or objective measures) and user-driven (or subjective measures) are discussed and then analysed for their pros and cons. With the proposed new methodology, data-driven measures can be used to select some potentially interesting rules for the user's evaluation and the rules and the knowledge obtained during the evaluation can be employed to calculate user-driven measures for identifying interesting rules.

Julien Blanchard, Fabrice Guillet, and Pascale Juntz present a semantics-based classification of rule interestingness measures in **Chapter IV**. They propose a novel and useful classification of interestingness measures according to three criteria: the subject, the scope, and the nature of the measure. These criteria are essential to grasp the meaning of the measures, and therefore to help the users to choose the ones he/she wants to apply. Moreover, the classification allows one to compare the rules to closely related concepts such as similarities, implications, and equivalences.

Section III presents four techniques on post-analysis and post-mining of association rules. **Chapter V** by Huawen Liu, Jigui Sun, and Huijie Zhang presents a technique on post-processing for rule reduction using closed set. Superfluous rules are filtered out from knowledge base in a post-processing manner. With dependent relation discovered by closed set mining technique, redundant rules can be eliminated efficiently.

In **Chapter VI**, Hacène Cherfi, Amedeo Napoli, and Yannick Toussaint present a new technique to combine data mining and semantic techniques for post-mining and selection of association rules. To focus on the result interpretation and discover new knowledge units, they introduce an original approach to classify association rules according to qualitative criteria using domain model as background knowledge. Its successful application on text mining in molecular biology shows the benefits of taking into account a knowledge domain model of the data.

In the case of stream data, the post-mining of association is more challenging. **Chapter VII** by Hetal Thakkar, Barzan Mozafari, and Carlo Zaniolo present a technique for continuous post-mining of association rules in a data stream management system. The chapter describes the architecture and techniques used to achieve this advanced functionality in the Stream Mill Miner (SMM) prototype, an SQL-based DSMS designed to support continuous mining queries.

The Receiver Operating Characteristics (ROC) graph is a popular way of assessing the performance of classification rules, but they are inappropriate to evaluate the quality of association rules, as there is no class in association rule mining and the consequent part of two different association rules might not have any correlation at all. Ronaldo Cristinano Prati presents in **Chapter VIII** a novel technique of QROC, a variation of ROC space to analyse itemset costs/benefits in association rules. It can be used to help analysts to evaluate the relative interestingness among different association rules in different cost scenarios.

Section IV presents rule selection techniques for classification. **Chapter IX** by Maria-Luiza Antonie, David Chodos, and Osmar Zaïane presents the rule generation, pruning, and selection in associative classifier, which is a classification model based on association rules. Several variations on the associative classifier model are presented, which are mining data sets with re-occurring items, using negative

association rules, and pruning rules using graph-based techniques. They also present a system, ARC-UI, which allows a user to analyze the results of classifying an item using an associative classifier.

In **Chapter X**, Silvia Chiusano and Paolo Garza discuss the selection of high-quality rules in associative classification. They present a comparative study of five well-known classification rule pruning methods and analyze the characteristics of both the selected and pruned rule sets in terms of information content. A large set of experiments has been run to empirically evaluate the effect of the pruning methods when applied individually as well as when combined.

Section V presents the visualization and representation techniques for the presentation and exploration of association rules. In **Chapter XI**, Sadok Ben Yahia, Olivier Couturier, Tarek Hamrouni, and Engelbert Mephu Nguifo, present two meta-knowledge based approaches for an interactive visualization of large amounts of association rules. Different from traditional methods of association rule visualization where association rule extraction and visualization are treated separately in a one-way process, the two proposed approaches that use meta-knowledge to guide the user during the mining process in an integrated framework covering both steps of the data mining process. The first one builds a roadmap of compact representation of association rules from which the user can explore generic bases of association rules and derive, if desired, redundant ones without information loss. The second approach clusters the set of association rules or its generic bases, and uses a fisheye view technique to help the user during the mining of association rules.

Chapter XII by Claudio Haruo Yamamoto, Maria Cristina Ferreira de Oliveira, and Solange Oliveira Rezende also discusses the visualization techniques to assist the generation and exploration of association rules. It presents an overview of the many approaches on using visual representations and information visualization techniques to assist association rule mining. A classification of the different approaches that rely on visual representations is introduced, based on the role played by the visualization technique in the exploration of rule sets. A methodology that supports visually assisted selective generation of association rules based on identifying clusters of similar itemsets is also presented. Then, a case study and some trends/issues for further developments are presented.

Nicolas Pasquier presents in **Chapter XIII** frequent closed itemset based condensed representations for association rules. Many applications of association rules to data from different domains have shown that techniques for filtering irrelevant and useless association rules are required to simplify their interpretation by the end-user. This chapter focuses on condensed representations that are characterized in the frequent closed itemsets framework to expose their advantages and drawbacks.

Section VI present techniques on the maintenance of association rules and new forms of association rules. **Chapter XIV** by Mengling Feng, Jinyan Li, Guozhu Dong, and Limsoon Wong presents a survey of the techniques for the maintenance of frequent patterns. The frequent pattern maintenance problem is summarized with a study on how the space of frequent patterns evolves in response to data updates. Focusing on incremental and decremental maintenance, four major types of maintenance algorithms are introduced, and the advantages and limitations of these algorithms are studied from both the theoretical and experimental perspectives. Possible solutions to certain limitations and potential research opportunities and emerging trends in frequent pattern maintenance are also discussed.

Conditional contrast patterns are designed by Guozhu Dong, Jinyan Li, Guimei Liu, and Limsoon Wong in **Chapter XV**. It is related to contrast mining, where one considers the mining of patterns/ models that contrast two or more datasets, classes, conditions, time periods, etc. Roughly speaking, conditional contrasts capture situations where a small change in patterns is associated with a big change in the matching data of the patterns. It offers insights on "discriminating" patterns for a given condition. It can also be viewed as a new direction for the analysis and mining of frequent patterns. The chapter formalizes the concepts of conditional contrast and provides theoretical results on conditional contrast mining. An efficient algorithm is proposed based on the results and experiment results are reported.

In **Chapter XVI**, Qinrong Feng, Duoqian Miao, and Ruizhi Wang present a technique for multidimensional model-based decision rules mining, which can output generalized rules with different degree of generalization. A method of decision rules mining from different abstract levels is provided in the chapter, which aims to improve the efficiency of decision rules mining by combining the hierarchical structure of multidimensional model and the techniques of rough set theory.

IMPACTS AND CONTRIBUTIONS

By collecting the research on the post-mining, summarization and presentation of association rule, as well as new forms and trends of association rules, this book shows the advanced techniques for the post-processing stage of association rules and presents readers what can be done to extract useful and actionable knowledge after discovering a large number of association rules. It will foster the research in the above topic and will benefit the use of association rule mining in real world applications. The readers can develop a clear picture on what can be done after discovering many association rules to extract useful knowledge and actionable patterns. Readers from industry can benefit by discovering how to deal with the large number of rules discovered and how to summarize or visualize the discovered rules to make them applicable in business applications. As editors, we hope this book will encourage more research into this area, stimulate new ideas on the related topics, and lead to implementations of the presented techniques in real-world applications.