Preface

This is the third volume of the Advances in Data Warehousing and Mining (ADWM) book series. ADWM publishes books in the areas of data warehousing and mining. This special volume, *Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection*, presents cutting edge research in this newly emerging area. Techniques for rare association mining are quite different from that of traditional rule mining and this book fills an essential gap in this area.

The primary objective of this book is to give readers in-depth knowledge on the current issues in rare association rule mining and critical event detection. The book is designed to cover a comprehensive range of topics related to rare association rule mining and critical event detection: mining techniques, imbalanced datasets, interest metrics, and real-world application domains. We hope this book will highlight the need for growth and research in the area of rare association rule mining and critical event detection. This volume consists of sixteen chapters in four sections.

The first section, Beyond the Support-Confidence Framework, provides an introduction to the area of rare association rule mining, and looks at some of the current proposed techniques which have moved away from the traditional support and confidence measures. This section contains four chapters.

Chapter 1, "Rare Association Rule Mining: Overview", by Yun Sing Koh, Auckland University of Technology, New Zealand, and Nathan Rountree, University of Otago, New Zealand, introduces the problem faced in the area of rare association rule mining and the current trends in this area. They provide an extensive literature review on the currently available techniques when dealing with rare itemsets.

Chapter 2, "Association Rule and Quantitative Association Rule Mining among Infrequent Items", by Ling Zhou and Stephen Yau, University of Illinois at Chicago, proposes two new methods to mine infrequent items and find rare association rules. Their approach is versatile and can also be applied to frequent items with bounded length. In addition they explore quantitative association rule mining among infrequent items by associating quantities of items: some interesting examples are drawn to illustrate the significance of such mining.

Chapter 3, "Replacing Support in Association Rule Mining", by Rosa Meo and Dino Ienco, Università di Torino, Italy, proposes a new model which adopts criteria based on Bayes’ Theorem and on an estimate of the probability density function of each itemset to establish the reliability of the information extracted from the database.

Chapter 4, "Effective Mining of Weighted Fuzzy Association Rules", by Maybin Muyeba, Manchester Metropolitan University, UK, M. Sulaiman Khan, Liverpool Hope University, UK, and Frans Coenen, University of Liverpool, UK, presents a novel approach for effectively mining weighted fuzzy association rules. They generalize the weighted association rule mining problem with binary and fuzzy attributes with weighted settings.

The second section, Dealing with Imbalanced Datasets, looks at algorithms and mining frameworks for dealing with datasets where there is uneven representation of various database objects. Imbalanced data
is a key issue in rare association rule mining, because: a) it is a necessary condition of rare itemsets, and 
b) it affects the power and accuracy of the statistical models used to perform data mining. This section 
consists of three chapters, where we look at rare class association rule mining, sub-class association rule 
mining, and mining minimal infrequent elements.

Chapter 5, “Rare Class Association Rule Mining with Multiple Imbalanced Attributes”, by Huaifeng 
Zhang, Yanchang Zhao, Longbing Cao, Chengqi Zhang, University of Technology, Sydney, Australia, and 
Hans Bohnescheid, Projects Section, Business Integrity Programs Branch, Centrelink, Australia, proposes 
a framework for rare class association rule mining. In their approach, the rules without imbalanced 
attributes are mined through a standard algorithm while the rules with imbalanced attributes are mined 
based on newly defined measurements. In this chapter, they present a compelling case study applied in 
the social security field.

Chapter 6, “A Multi-Methodological Approach to Rare Association Rule Mining” by Yun Sing Koh, 
Auckland University of Technology, New Zealand, and Russell Pears, Auckland University of Technology, 
New Zealand, proposes a synthesis of material from three different methodologies to tackle the problem 
of rare association rule mining: itemset weighting, clustering, and statistical significance testing. They 
focus on the importance of sub-class rare rules or absolute rare rules. Absolute rare rules are those that 
are not just rare to the dataset as a whole but are also rare to the cluster from which they are derived.

Chapter 7, “Finding Minimal Infrequent Elements in Multi-Dimensional Data Defined over Partially 
Ordered Sets and its Applications”, by Khaled M. Elbassi, Max-Planck-Institut fur Informatik, 
Germany, studies the complexity of finding all minimal infrequent elements for some interesting classes 
of partially ordered set (poset). He looks at a general framework used to mine associations from different 
types of databases. The rules obtained under this framework are generally stronger than the ones obtained 
from techniques that use binarization.

In Section 3, Rare, Anomalous, and Interesting Patterns, we look at some of the techniques used to 
find interesting and unexpected patterns in the area of association rules. Section three consists of five 
chapters, discussing issues related to discovering interesting patterns in numerical data with background 
knowledge, discovering quasi-functional dependencies, mining unexpected patterns, and extracting 
anomalous rules.

Chapter 8, “Discovering Interesting Patterns in Numerical Data with Background Knowledge”, by Szymon Jaroszewicz, National Institute of Telecommunications, Warsaw, Poland, presents an approach 
to mining patterns in numerical data without the need for discretization. The proposed method allows 
for discovery of arbitrary nonlinear relationships where the user may include background knowledge in 
the form of a probabilistic model. The patterns that have been previously predicted by the model will 
not be considered interesting. Interesting patterns can then be used by the user to update the probabilistic 
model.

Chapter 9, “Mining Rare Association Rules by Discovering Quasi-functional Dependencies: An 
Incremental Approach”, by Giulia Bruno and Paolo Garza, Politecnico di Torino Corso Duca degli Abruzzi, 
Italy, and Elisa Quintarell, Politecnico di Milano Piazza Leonardo da Vinci, Italy, propose a method of 
detecting rare rules by first inferring the normal behaviour of objects in the form of quasi-functional 
dependencies (i.e. functional dependencies that frequently hold), and then analysing rare violations with 
respect to them. They propose an incremental algorithm to efficiently maintain up-to-date rules.

Chapter 10, “Mining Unexpected Sequential Patterns and Implication Rules” by Dong (Haoyuan) 
Li, LGI2P, École des Mines d’Alès, France, Anne Laurent and Pascal Poncelet, LIRMM, Université 
Montpellier II, France, presents an approach called USER for mining unexpected sequential rules in 
sequence databases. They propose a belief-driven formalization of the unexpectedness contained in 
sequential data.
Chapter 11, “Mining Hidden Association Rules from Real-Life Data” by Marco-Antonio Balderas Cepeda, Universidad de Granada, Spain, provides an adaptation of measures of interest to our anomalous rule sets, and proposed an algorithm that can extract anomalous rules as well. Their approach discovered hidden patterns with good reliability.

Chapter 12, “Strong Symmetric Association Rules and Interestingness Measures” by Agathe Merceron, University of Applied Sciences TFH Berlin, Germany, proposes a method to find strong symmetric association rules. This approach is slightly different from the conventional rare association rule mining. This kind of rule can be qualified as rare, as they would be pruned by many objective interestingness measures.

In Section 4, Critical Event Detection and Applications, we look at some of the applications of rare association rule mining and critical event detection. In this section, we provide two chapters which specifically look at the usage of association rule mining in different domains. The last two chapters look at a different data mining approach, namely classification techniques, for critical event detection. The areas of application discussed include adverse drug reaction monitoring, analysis of traffic accident, risk levels for violent felony crimes, and financial credit monitoring.

Chapter 13, “He Wasn’t There Again Today”, by Richard O’Keefe and Nathan Rountree, University of Otago, New Zealand, discusses the characteristics of data collected by the New Zealand Centre for Adverse Drug Reaction Monitoring (CARM) over a five-year period. They discuss the notion of “rarity” with respect to drugs, and with respect to reactions.

Chapter 14, “Filtering Association Rules by Their Semantics and Structures” by Rangsipan Marukatat, Mahidol University, Thailand, introduces the filtering of association rules by their patterns and degrees of semantic redundancy. They applied their techniques to a real case study, an analysis of traffic accidents in Nakorn Pathom, Thailand.

Chapter 15, “Creating Risk-Scores in very Imbalanced Datasets: Predicting Extremely Violent Crime among Criminal Offenders Following Release from Prison” by Markus Breitenbach, William Dieterich, Tim Brennan, Northpointe Institute for Public Management, USA, and Adrian Fan, University of Colorado at Boulder, USA, explores the Area under Curve (AUC) as an error-metric suitable for imbalanced data, as well as survey methods of optimizing this metric directly. They conducted a study that examines predictive rule development and validation procedures for establishing risk levels for violent felony crimes committed when criminal offenders are released from prison in the USA.

Chapter 16, “Boosting Prediction Accuracy of Bad Payments in Financial Credit Applications”, by Russel Pears and Raymond Oetama, Auckland University of Technology, New Zealand, use a machine learning approach to improve the identification of such customers. They proposed a credit scoring approach to predict bad payments for credit risk management.

We hope that this book will provide readers some specific challenge that motivates the development and enhancement of rare association rule mining and critical event detection area. We also hope that this book will serve as an introductory material to the researchers and practitioners interested in this emerging area of research.

Yun Sing Koh and Nathan Rountree
January 2009