

Computational Risk Management

Editors-in-Chief

Desheng Dash Wu, RiskLab, University of Toronto, Toronto, ON, Canada

David L. Olson, Department of Supply Chain Management and Analytics,
University of Nebraska-Lincoln, Lincoln, NE, USA

John Birge, University of Chicago Booth School of Business, Chicago, IL, USA

Risks exist in every aspect of our lives and risk management has always been a vital topic. Most computational techniques and tools have been used for optimizing risk management and the risk management tools benefit from computational approaches. Computational intelligence models such as neural networks and support vector machines have been widely used for early warning of company bankruptcy and credit risk rating. Operational research approaches such as VaR (value at risk) optimization have been standardized in managing markets and credit risk, agent-based theories are employed in supply chain risk management and various simulation techniques are employed by researchers working on problems of environmental risk management and disaster risk management. Investigation of computational tools in risk management is beneficial to both practitioners and researchers. The Computational Risk Management series is a high-quality research book series with an emphasis on computational aspects of risk management and analysis. In this series, research monographs as well as conference proceedings are published.

More information about this series at <http://www.springer.com/series/8827>

David L. Olson · Georg Lauhoff

Descriptive Data Mining

Second Edition

 Springer

David L. Olson
College of Business
University of Nebraska–Lincoln
Lincoln, NE, USA

Georg Lauhoff
San Jose, CA, USA

ISSN 2191-1436 ISSN 2191-1444 (electronic)
Computational Risk Management
ISBN 978-981-13-7180-6 ISBN 978-981-13-7181-3 (eBook)
<https://doi.org/10.1007/978-981-13-7181-3>

Library of Congress Control Number: 2019934798

© Springer Nature Singapore Pte Ltd. 2017, 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Knowledge management involves the application of human knowledge (epistemology) with the technological advances of our current society (computer systems) and big data, both in terms of collecting data and in analyzing it. We see three types of analytic tools. **Descriptive** analytics focus on the reports of what has happened. **Predictive** analytics extend statistical and/or artificial intelligence to provide forecasting capability. It also includes classification modeling. **Diagnostic** analytics can apply analysis to sensor input to direct control systems automatically. **Prescriptive** analytics applies quantitative models to optimize systems, or at least to identify improved systems. Data mining includes descriptive and predictive modeling. Operations research includes all the three. This book focuses on descriptive analytics.

Lincoln, USA
San Jose, USA

David L. Olson
Georg Lauhoff

Book Concept

The book seeks to provide simple explanations and demonstration of some descriptive tools. This second edition provides more examples of big data impact, updates the content on visualization, clarifies some points, and expands coverage of association rules and cluster analysis. Chapter 1 gives an overview of the context of knowledge management. Chapter 2 discusses some basic software support to data visualization. Chapter 3 covers fundamentals of market basket analysis, and Chap. 4 provides a demonstration of RFM modeling, a basic marketing data mining tool. Chapter 5 demonstrates association rule mining. Chapter 6 has more in-depth coverage of cluster analysis. Chapter 7 discusses link analysis.

Models are demonstrated using business-related data. The style of the book is intended to be descriptive, seeking to explain how methods work, with some citations, but without deep scholarly references. The data sets and software are all selected for widespread availability and access by any reader with computer links.

Contents

- 1 Knowledge Management** 1
 - Computer Support Systems 2
 - Examples of Knowledge Management 4
 - Data Mining Descriptive Applications 7
 - Summary 8
 - References 8
- 2 Data Visualization** 11
 - Data Visualization 11
 - R Software 12
 - Loan Data 13
 - Energy Data 20
 - Basic Visualization of Time Series 21
 - Conclusion 28
 - References 30
- 3 Market Basket Analysis** 31
 - Definitions 32
 - Co-occurrence 33
 - Demonstration 37
 - Fit 38
 - Profit 38
 - Lift 41
 - Market Basket Limitations 43
 - References 44
- 4 Recency Frequency and Monetary Analysis** 45
 - Dataset 1 46
 - Balancing Cells 50
 - Lift 52
 - Value Function 53

Data Mining Classification Models	58
Logistic Regression	58
Decision Tree	59
Neural Networks	59
Dataset 2	59
Conclusions	63
References	65
5 Association Rules	67
Methodology	68
The Apriori Algorithm	69
Association Rules from Software	71
Non-negative Matrix Factorization	75
Conclusion	76
References	76
6 Cluster Analysis	77
K-Means Clustering	78
A Clustering Algorithm	78
Loan Data	79
Clustering Methods Used in Software	81
Software	82
R (Rattle) K-Means Clustering	82
Other R Clustering Algorithms	88
KNIME	96
WEKA	98
Summary	105
References	106
7 Link Analysis	107
Link Analysis Terms	107
Basic Network Graphics with NodeXL	114
Network Analysis of Facebook Network or Other Networks	118
Link Analysis of Your Emails	124
Link Analysis Application with PolyAnalyst (Olson and Shi 2007)	125
Summary	128
References	128
8 Descriptive Data Mining	129

About the Authors

David L. Olson is the James & H.K. Stuart Chancellor's Distinguished Chair and Full Professor at the University of Nebraska. He has published research in over 150 refereed journal articles, primarily on the topic of multiple objective decision-making, information technology, supply chain risk management, and data mining. He teaches in the management information systems, management science, and operations management areas. He has authored over 20 books. He is Member of the Decision Sciences Institute, the Institute for Operations Research and Management Sciences, and the Multiple Criteria Decision Making Society. He was a Lowry Mays endowed Professor at Texas A&M University from 1999 to 2001. He was named the Raymond E. Miles Distinguished Scholar award for 2002, and was a James C. and Rhonda Seacrest Fellow from 2005 to 2006. He was named Best Enterprise Information Systems Educator by IFIP in 2006. He is a Fellow of the Decision Sciences Institute.

Georg Lauhoff is Technologist at Western Digital Corporation and carries out R&D in materials science and its application in data storage devices and uses the techniques described in this book for his work. He co-authored 38 refereed journal articles and over 30 conference presentations, primarily on the topic of materials science, data storage materials, and magnetic thin films. He was awarded scholarships and research grants in the UK and Japan. He was the Clerk Maxwell Scholar from 1995 to 1998 and is a Fellow of the Cambridge Philosophical Society. He studied physics at Aachen (Diplom) and Cambridge University (Master and Ph.D.) specializing in the field of materials science and magnetic thin films and sensors. After graduating, he moved to Japan and held a faculty position in Materials Science and Engineering at the Toyota Technological Institute and then carried out research in the sequencing of DNA using magnetic sensors at Cambridge University before moving in 2005 to the recording industry in the Bay area.