SPRINGER BRIEFS IN BUSINESS

Yong Shi
Lingling Zhang
Yingjie Tian
Xingsen Li

Intelligent Knowledge

A Study beyond Data Mining

Springer

SPRINGER BRIEFS IN BUSINESS

Yong Shi
Lingling Zhang
Yingjie Tian
Xingsen Li

Intelligent Knowledge

A Study beyond Data Mining

Springer

SpringerBriefs in Business

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic. Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

SpringerBriefs in Business showcase emerging theory, empirical research, and practical application in management, finance, entrepreneurship, marketing, operations research, and related fields, from a global author community.

Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, standardized manuscript preparation and formatting guidelines, and expedited production schedules.

More information about this series at http://www.springer.com/series/8860

Yong Shi • Lingling Zhang • Yingjie Tian
Xingsen Li

Intelligent Knowledge

A Study Beyond Data Mining

Springer

Yong Shi
Research Center on Fictitious Economy and Data Science
Chinese Academy of Sciences
Beijing
China

Yingjie Tian
Research Center on Fictitious Economy and Data Science
Chinese Academy of Sciences
Beijing
China

Lingling Zhang
School of Management
University of Chinese Academy of Sciences
Beijing
China

Xingsen Li
School of Management,
Ningbo Institute of Technology, Zhejiang University
Ningbo
Zhejiang
China

ISSN 2191-5482 ISSN 2191-5490 (electronic)
SpringerBriefs in Business
ISBN 978-3-662-46192-1 ISBN 978-3-662-46193-8 (eBook)
DOI 10.1007/978-3-662-46193-8

Library of Congress Control Number: 2014960237

Springer Berlin Heidelberg New York Dordrecht London

Printed on acid-free paper

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

To all of Our Colleagues and Students at Chinese Academy of Sciences

Preface

This book provides a fundamental method of bridging data mining and knowledge management, which are two important fields recognized respectively by the information technology (IT) community and business analytics (BA) community. For a quit long time, IT community agrees that the results of data mining are "hidden patterns", not "knowledge" yet for the decision makers. In contrast, BA community needs the explicit knowledge from large database, now called Big Data in addition to implicit knowledge from the decision makers. How to human experts can incorporate their experience with the knowledge from data mining for effective decision support is a challenge. There some previous research on post data mining and domain-driven data mining to address this problem. However, the findings of such researches are preliminary; either based on heuristic learning, or experimental studies. They have no solid theoretical foundations. This book tries to answer the problem by a term, called "Intelligent Knowledge."

The motivation of the research on Intelligent Knowledge was started with a business project carried out by the authors in 2006 (Shi and Li, 2007). NetEase, Inc., a leading China-based Internet technology company, wanted to reduce its serious churn rate from the VIP customers. The customers can be classified as "current users, freezing users and lost users". Using a well-known tool of decision tree classification algorithm, the authors found 245 rules from thousands of rules, which could not tell the knowledge of predicting user types. When the results were presented to a marketing manager of the company, she, with her working experience (domain knowledge), immediately selected a few rules (decision support) from 245 results. She said, without data mining, it is impossible to identify the rules to be used as decision support. It is data mining to help her find 245 hidden patterns, and then it is her experience to further recognize the right rules. This lesson trigged us that the human knowledge must be applied on the hidden patterns from data mining. The research is to explore how human knowledge can be systematically used to scan the hidden patterns so that the latter can be upgraded as the "knowledge" for decision making. Such "knowledge" in this book is defined as Intelligent Knowledge.

When we proposed this idea to the National Science Foundation of China (NSFC) in the same year, it generously provided us its most prestigious fund, called

"the Innovative Grant" for 6 years (2007–2012). The research findings presented in this book is part of the project from NSFC's grant as well as other funds.

Chapter 1–6 of this book is related to concepts and foundations of Intelligent Knowledge. Chapter 1 reviews the trend of research on data mining and knowledge management, which are the basis for us to develop intelligent knowledge. Chapter 2 is the key component of this book. It establishes a foundation of intelligent knowledge management over large databases or Big Data. Intelligent Knowledge is generated from hidden patterns (it then called "rough knowledge" in the book) incorporated with specific, empirical, common sense and situational knowledge, by using a "second-order" analytic process. It not only goes beyond the traditional data mining, but also becomes a critical step to build an innovative process of intelligent knowledge management—a new proposition from original data, rough knowledge, intelligent knowledge, and actionable knowledge, which brings a revolution of knowledge management based on Big Data. Chapter 3 enhances the understanding about why the results of data mining should be further analyzed by the second-order data mining. Through a known theory of Habitual Domain analysis, it examines the effect of human cognition on the creation of intelligent knowledge during the second-order data mining process. The chapter shows that people's judgments on different data mining classifiers diverge or converge can inform the design of the guidance for selecting appropriate people to evaluate/select data mining models for a particular problem. Chapter 4 proposes a framework of domain driven intelligent knowledge discovery and demonstrate this with an entire discovery process which is incorporated with domain knowledge in every step. Although the domain driven approaches have been studied before, this chapter adapts it into the context of intelligent knowledge management to using various measurements of interestingness to judge the possible intelligent knowledge. Chapter 5 discusses how to combine prior knowledge, which can be formulated as mathematical constraints, with well-known approaches of Multiple Criteria Linear Programming (MCLP) to increase possibility of finding intelligent knowledge for decision makers. The proposed is particular important if the results of a standard data mining algorithm cannot be accepted by the decision maker and his or her prior (domain) knowledge can be represented as mathematical forms. Following the similar idea of Chapter 5, when the human judgment can expressed by certain rules, then Chapter 6 provides a new method to extract knowledge, with a thought inspired by the decision tree algorithm, and give a formula to find the optimal attributes for rule extraction. This chapter demonstrates how to combine different data mining algorithms (Support vector Machine and decision tree) with the representation of human knowledge in terms of rules.

Chapter 7–8 of this book is about the basic applications of Intelligent Knowledge. Chapter 7 elaborates a real-life intelligent knowledge management project to deal with customer churn in NetEase, Inc.. Almost all of the entrepreneurs desire to have brain trust generated decision to support strategy which is regarded as the most critical factor since ancient times. With the coming of economic globalization era, followed by increasing competition, rapid technological change as well as gradually accrued scope of the strategy. The complexity of the explosive increase made only by the human brain generates policy decision-making appeared to be inadequate.

Chapter 8 applies a semantics-based improvement of Apriori algorithm, which integrates domain knowledge to mining and its application in traditional Chinese Medicines. The algorithm can recognize the changes of domain knowledge and re-mining. That is to say, the engineers need not to take part in the course, which can realize intellective acquirement.

This book is dedicated to all of our colleagues and students at the Chinese Academy of Sciences. Particularly, we are grateful to these colleagues who have working with us for this meaningful project: Dr. Yinhua Li (China Merchants Bank, China), Dr. Zhengxiang Zhu (the PLA National Defense University, China), Le Yang (the State University of New York at Buffalo, USA), Ye Wang (National Institute of Education Sciences, China), Dr. Guangli Nie (Agricultural Bank of China, China), Dr. Yuejin Zhang (Central University of Finance and Economics, China), Dr. Jun Li (ACE Tempest Reinsurance Limited, China), Dr. Bo Wang (Chinese Academy of Sciences), Mr. Anqiang Huang (BeiHang University, China), Zhongbiao Xiang(Zhejiang University, China)and Dr. Quan Chen (Industrial and Commercial Bank of China, China). We also thank our current graduate students at Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences: Zhensong Chen, Xi Zhao, Yibing Chen, Xuchan Ju, Meng Fan and Qin Zhang for their various assistances in the research project.

Finally, we would like acknowledge a number of funding agencies who supported our research activities on this book. They are the National Natural Science Foundation of China for the key project “Optimization and Data Mining,” (#70531040, 2006–2009), the innovative group grant “Data Mining and Intelligent Knowledge Management,” (#70621001, #70921061, 2007–2012); Nebraska EPScOR, the National Science Foundation of USA for industrial partnership fund “Creating Knowledge for Business Intelligence” (2009–2010); Nebraska Furniture Market—a unit of Berkshire Hathaway Investment Co., Omaha, USA for the research fund “Revolving Charge Accounts Receivable Retrospective Analysis,” (2008–2009); the CAS/SAFEA International Partnership Program for Creative Research Teams “Data Science-based Fictitious Economy and Environmental Policy Research” (2010–2012); Sojern, Inc., USA for a Big Data research on “Data Mining and Business Intelligence in Internet Advertisements” (2012–2013); the National Natural Science Foundation of China for the project “Research on Domain Driven Second Order Knowledge Discovering” (#71071151, 2011–2013); National Science Foundation of China for the international collaboration grant “Business Intelligence Methods Based on Optimization Data Mining with Applications of Financial and Banking Management” (#71110107026, 2012–2016); the National Science Foundation of China, Key Project “Innovative Research on Management Decision Making under Big Data Environment” (#71331005, 2014–2018); the National Science Foundation of China, “Research on mechanism of the intelligent knowledge emergence of innovation based on Extenics” (#71271191, 2013–2016) the National Natural Science Foundation of China for the project “Knowledge Driven Support Vector Machines Theory, Algorithms and Applications” (#11271361, 2013–2016) and the National Science Foundation of China. “The Research of Personalized Recommend System Based on Domain Knowledge and Link Prediction” (#71471169,2015-2018).

Contents

About the Authors

Yong Shi serves as the Executive Deputy Director, Chinese Academy of Sciences Research Center on Fictitious Economy & Data Science. He is the Union Pacific Chair of Information Science and Technology, College of Information Science and Technology, Peter Kiewit Institute, University of Nebraska, USA. Dr. Shi's research interests include business intelligence, data mining, and multiple criteria decision making. He has published more than 20 books, over 200 papers in various journals and numerous conferences/proceedings papers. He is the Editor-in-Chief of International Journal of Information Technology and Decision Making (SCI), Editor-in-Chief of Annals of Data Science (Springer), and a member of Editorial Board for a number of academic journals. Dr. Shi has received many distinguished awards including the Georg Cantor Award of the International Society on Multiple Criteria Decision Making (MCDM), 2009; Fudan Prize of Distinguished Contribution in Management, Fudan Premium Fund of Management, China, 2009; Outstanding Young Scientist Award, National Natural Science Foundation of China, 2001; and Speaker of Distinguished Visitors Program (DVP) for 1997-2000, IEEE Computer Society. He has consulted or worked on business projects for a number of international companies in data mining and knowledge management.

Lingling Zhang received her PhD from Bei Hang University in 2002. She is an Associate Professor at University of Chinese Academy of Sciences since 2005. She also works as a Researcher Professor at Research Center on Fictitious Economy and Data Science and teaches in Management School of University of Chinese Academy of Sciences. She has been a visiting scholar of Stanford University. Currently her research interest covers intelligent knowledge management, data mining, and management information system. She has received two grant supported by the Natural Science Foundation of China (NSFC), published 4 books, more than 50 papers in various journals and some of them received good comments from the academic community and industries.

Yingjie Tian received the M.Sc. degree from Beijing Institute of Technology, in 1999, and the Ph.D. degree from China Agricultural University, Beijing, China, in 2005. He is currently a Professor with the Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, China. He has authored

four books about support vector machines, one of which has been cited over 1000 times. His current research interests include support vector machines, optimization theory and applications, data mining, intelligent knowledge management, and risk management.

Xingsen Li received the M.Sc degree from China University of Mining and Technology Beijing in 2000, and the Ph.D. degree in management science and engineering from Graduate University of Chinese Academy of Sciences in 2008. He is currently a Professor in NIT, Zhejiang University and a director of Chinese Association for Artificial Intelligence (CAAI) and the Secretary-General of Extension engineering committee, CAAI. He has authored two books about intelligent knowledge management and Exteincs based data mining. His current research interests include intelligent knowledge management, big data, Extenics-based data mining and Extenics-based innovation.