

Data Mining and Multi-agent Integration

Longbing Cao
Editor

Data Mining and Multi-agent Integration

 Springer

Editor

Longbing Cao
Faculty of Engineering and Information Technology
University of Technology, Sydney
Broadway, NSW 2007
Australia
lbcao@it.uts.edu.au

ISBN 978-1-4419-0521-5 e-ISBN 978-1-4419-0522-2
DOI 10.1007/978-1-4419-0522-2
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009928846

© Springer Science+Business Media, LLC 2009

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

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

Preface

Data Mining and Multi-agent Integration aims to reflect state-of-the-art research and development of *agent mining interaction and integration* (for short, agent mining).

The book was motivated by increasing interest and work in the agents data mining, and vice versa. The interaction and integration comes about from the intrinsic challenges faced by agent technology and data mining respectively; for instance, multi-agent systems face the problem of enhancing agent learning capability, and avoiding the uncertainty of self-organization and intelligence emergence. Data mining, if integrated into agent systems, can greatly enhance the learning skills of agents, and assist agents with predication of future states, thus initiating follow-up action or intervention. The data mining community is now struggling with mining distributed, interactive and heterogeneous data sources. Agents can be used to manage such data sources for data access, monitoring, integration, and pattern merging from the infrastructure, gateway, message passing and pattern delivery perspectives. These two examples illustrate the potential of agent mining in handling challenges in respective communities.

There is an excellent opportunity to create innovative, dual agent mining interaction and integration technology, tools and systems which will deliver results in one new technology. For example, if an open complex agent system is powered with actionable knowledge discovery capabilities, it then has the potential to deal with very complex problem solving with super-intelligent information processing, knowledge discovery, collective intelligence emergence, and actionable decision-making skills in complex environments. Currently, systems of this magnitude are not possible without the integration of agents and data mining.

This book, as the first in this area, does not intend to cover the field of agent mining. Rather, it features the latest methodological, technical and practical progress on promoting the successful use of agent mining. In 22 chapters, the book reflects state-of-the-art agent mining research and development. The book is divided into three parts. Part I provides an introduction to agents and data mining integration. Part II addresses data mining-driven agents, and Part III focuses on agent-driven data mining.

Part I has three introductory chapters. Chapter One presents a comprehensive introduction to interaction and integration of agents and data mining, which covers driving forces, disciplinary frameworks, agent-driven distributed data mining, data mining driven agents, mutual issues in agent mining, applications and case studies, trends and directions, and agent-mining community development. Chapter Two presents a brief overview of agent mining interaction, and two case studies. Chapter Three provides a survey on agent-based distributed data mining.

Part II has nine chapters outlining the latest progress and techniques for data mining driven agents. Chapter Four explores agent behavior by particle swarm optimization based web usage clustering. Chapter Five enhances agent learning through mining temporal patterns of agent behavior. Chapter Six summarizes the use of Web and structure mining to form an agent system which detects user interaction information. Chapter Seven presents an e-commerce-oriented distributed recommender system with peer profiling and selection strategy. Chapter Eight uses multi-class classification to build a multi-agent-based intrusion detection system. Chapter Nine proposes genetic algorithms and regular expressions to automatically learn about software entities. Chapter 10 proposes a words weight vectors driven network module for establishing and maintaining a knowledge network. Chapter 11 proposes goal mining of query logs for commonsense knowledge to equip intelligent agents. Chapter 12 proposes an agent-based interactive diagnostic workbench with diagnostic rules.

In Part III, we present 10 chapters on agents-driven data mining. Chapter 13 proposes an extensible multi-agent data mining system powered by association rule and classification. Chapter 14 proposes an anytime multi-agent approach to on-line unsupervised learning, which handles continuous agglomerative hierarchical clustering of streaming data. Chapter 15 proposes an agent system utilizing a divide and conquer approach and data dependent schemes for clustering large data. Chapter 16 proposes a multi-ant colony and a multi-objective clustering algorithm by combining the results of all colonies. Chapter 17 proposes an interactive environment for psychometrics diagnostics, where agents supervise the users actions and data mining for searching of potentially interesting information. Chapter 18 uses static agent to execute mine firewall policy rules, and mobile agent to exploit optimized rules to detect eventual anomalies. Chapter 19 uses game-theory modeling for competitive knowledge extraction, hierarchical knowledge mining, and Dempster-Shafer result combination. Chapter 20 discusses a normative multi-agent enriched data mining architecture and ontology frameworks. Chapter 21 presents static and dynamic agent societies responsible for group formation, to execute a data mining classification process. Chapter 22 describes an agent based video contents identification scheme using a watermark based filtering technique.

This book is directed to students, researchers, engineers and practitioners in both the agent and data mining areas, who are interested in the marriage of agents with data mining, or are experiencing challenges in one area that could be managed by incorporating the other technology. The book will also be of interest to students and researchers in many related areas such as machine learning, artificial intelligence, intelligent systems, knowledge engineering, human-computer interaction, intelligent

information processing, decision support systems, knowledge management, organizational computing, social computing, complex systems, and soft computing.

We would like to convey our appreciation to all contributors, including the accepted chapters' authors, and many other participants who submitted work that has not been included in this book. Our special thanks go to Ms. Melissa Fearon and Ms. Valerie Schofield from Springer US for their kind support and efforts in bringing the book to fruition. In addition, we also appreciate our reviewers, and Ms. Ziyue Zuo and Mr. Yong Yang's assistance in co-ordinating the book chapter collection and forming the book.

Longbing Cao

University of Technology Sydney, Australia

March 2009

Foreword

Integration of advanced technologies is a natural progression in the development of science. Such integration may result in a new technology whose “power” can be metaphorically described by the equality $2 + 2 = 5$, or even more. The integration of agents and data mining is such an emergent technology. The potential benefits of integrating agent and data mining were recognized when multi-agent systems (MAS) was at its infancy stage and when data mining has already reached close to their maturity. Since that time research has demonstrated the value of integration.

These advantages can be evaluated from many viewpoints and perspectives. Indeed, enrichment of MAS with data mining capabilities can significantly enhance intelligence of multi-agent systems by enabling them to adapt to unpredictable changes of environment, and by providing them with a property to on-line improve agents’ collaboration utility through learning. The latter is especially important for novel classes of applications, in particular, for mobile, ubiquitous and peer-to-peer networking systems where high level of autonomy of agents pursuing their local goals has to be harmonized with the global system goal. Distributed data mining and learning technologies can provide such applications with powerful means for on-line coordination and agreement of local and global system objectives. An example is given by so-called agent network self-configuration task. Indeed, a specific feature of mobile, ubiquitous and peer-to-peer (P2P) computing systems is that they operate in dynamic environment. E.g., mobile devices may move and freely enter into and exit from a mobile network changing the set of network nodes and their communication topology and thus changing the set of application agents existing in the network and the availability of services as well. Consistent and coherent operation of such networks is organized using a “virtualization” idea implemented in terms of overlay computing. In such dynamic environment, an important task is self-configuration of overlay networks to support the optimality of agent network operation at the communication, infrastructural and application layers. This task cannot be effectively solved without intensive use of distributed data mining and P2P machine learning technologies. The list of applications in which multi-agent systems, data mining and machine learning have been combined to great effect is growing rapidly. This book contains carefully selected examples.

From the data mining and machine learning perspective, multi-agent systems leads to new architectures and to novel styles of software development. Modern

data mining and machine learning applications lay down new requirements for enabling information and software technologies that are used at design and implementation phases. In particular, these, often challenging, requirements are caused by the distributed nature and heterogeneity of data sources that are selected dynamically by a data mining application and depend on the current state of the environment. Data sources may be massive and may continue to generate formidable volumes of data. It is well recognized that, in such applications, the most promising technology is agent-based. It is capable of supporting multi-strategy data mining and of achieving scalability in the presence of massive quantities of distributed data. Agent technology can also support intelligent interaction with users. Experience shows that no other currently existing information technology can compare with the power of MAS technology for data mining problems in this class. However, this class of problems is not rare and includes modern real life data mining challenges.

Recently, researchers are intensively using agent technology in many novel data mining applications. Among them are: web mining, text mining, and brain computing. On the other side, agent-based applications use data mining and machine learning for coordination strategy mining, for interaction protocol selection and mining. A sign of the maturity of the integration of agent and data mining technologies is the development of several software tools specifically targeted for joint use of agent and data mining, for example, Agent Academy.

We acknowledge the important role of the FP5 European project KNet and, in particular, the AgentLink ALAD Special Interesting Group (SIG) organized within the framework of this Project, in consolidating the data mining and agent communities. ALAD SIG was a pioneer task force intended to achieve this consolidation leading to joint research. Another dedicated SIG is the Agent-Mining Interaction and Integration SIG¹, which actively links leading researchers in this emerging field, and organizes regular events for promoting and encouraging research networking in both agent and data mining communities. These task forces have put ahead many other analogous initiatives including special sessions of the international conferences, dedicated international workshops, journal special issues, etc.

As an explicit aftereffect of the AMII SIG initiative, this book on Data Mining and Multiagent Integration, edited by Longbing Cao, presents the first collection of the latest research outcomes on the topic. A single volume can not present a complete coverage of the current state-of-the-art in the perspectives of integrating agents and data mining, however it covers all major research topics of agent mining research, and presents many novel ideas, applications and other solutions that can be useful for the researchers and industry interested in research on and practical use of novel information technologies. This book will be of interest to new scientists working in agent and data mining integration and interaction, and will further promote the consolidation of this crucial international scientific community.

*Prof. Vladimir Gorodetsky,
St. Petersburg, SPIIRAS*

¹ AMII-SIG: www.agentmining.org

Contents

Part I Introduction to Agents and Data Mining Interaction

- 1 Introduction to Agent Mining Interaction and Integration 3**
Longbing Cao
- 2 Towards the Integration of Multiagent Applications and Data
Mining 37**
Célia Ghedini Ralha
- 3 Agent-Based Distributed Data Mining: A Survey 47**
Chayapol Moemeng, Vladimir Gorodetsky, Ziye Zuo, Yong Yang and
Chengqi Zhang

Part II Data Mining Driven Agents

- 4 Exploiting Swarm Behaviour of Simple Agents for Clustering Web
Users' Session Data 61**
Shafiq Alam, Gillian Dobbie, and Patricia Riddle
- 5 Mining Temporal Patterns to Improve Agents Behavior: Two Case
Studies 77**
Philippe Fournier-Viger, Roger Nkambou, Usef Fahihi, Engelbert
Mephu Nguifo
- 6 A Multi-Agent System for Extracting and Analysing Users'
Interaction in a Collaborative Knowledge Management System 93**
Doina Alexandra Dumitrescu, Ruth Cobos and Jaime Moreno-Llorena
- 7 Towards Information Enrichment through Recommendation
Sharing 103**
Li-Tung Weng, Yue Xu, Yuefeng Li and Richi Nayak

8	A Multiagent-based Intrusion Detection System with the Support of Multi-Class Supervised Classification	127
	Mei-Ling Shyu and Varsha Sainani	
9	Automatic Web Data Extraction Based on Genetic Algorithms and Regular Expressions	143
	David F. Barrero, David Camacho, and María D. R-Moreno	
10	Establishment and Maintenance of a Knowledge Network by Means of Agents and Implicit Data	155
	Jaime Moreno-Llorena, Xavier Alamán and Ruth Cobos	
11	Equipping Intelligent Agents with Commonsense Knowledge acquired from Search Query Logs: Results from an Exploratory Story	167
	Markus Strohmaier, Mark Kröll, and Peter Prettenhofer	
12	A Multi-Agent Learning Paradigm for Medical Data Mining Diagnostic Workbench	177
	Sam Chao and Fai Wong	

Part III Agent Driven Data Mining

13	The EMADS Extendible Multi-Agent Data Mining Framework	189
	Kamal Ali Albashiri, and Frans Coenen	
14	A Multiagent Approach to Adaptive Continuous Analysis of Streaming Data in Complex Uncertain Environments	201
	Igor Kiselev and Reda Alhajj	
15	Multiagent Systems for Large Data Clustering	219
	T.Ravindra Babu, M. Narasimha Murty, and S.V.Subrahmanya	
16	A Multiagent, Multiobjective Clustering Algorithm	239
	Daniela S. Santos, Denise de Oliveira, and Ana L. C. Bazzan	
17	Integration of Agents and Data Mining in Interactive Web Environment for Psychometric Diagnostics	251
	Velibor Ilić	
18	A Multi-Agent Framework for Anomalies Detection on Distributed Firewalls Using Data Mining Techniques	267
	Kamel Karoui, Fakher Ben Ftima and Henda Ben Ghezala	
19	Competitive-Cooperative Automated Reasoning from Distributed and Multiple Source of Data	279
	Amin Milani Fard	

20	Normative Multi-Agent Enriched Data Mining to Support E-Citizens	291
	Stanislaw A. B. Stane and Mariusz Zytniewsk	
21	CV-Muzar - The Virtual Community Environment that Uses Multiagent Systems for Formation of Groups	305
	Ana Carolina Bertoletti De Marchi and Márcia Cristina Moraes	
22	Agent based Video Contents Identification and Data Mining Using Watermark based Filtering	315
	HeungKyu Lee	
	Index	325