

Lecture Notes in Artificial Intelligence

10086

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel

University of Alberta, Edmonton, Canada

Yuzuru Tanaka

Hokkaido University, Sapporo, Japan

Wolfgang Wahlster

DFKI and Saarland University, Saarbrücken, Germany

LNAI Founding Series Editor

Joerg Siekmann

DFKI and Saarland University, Saarbrücken, Germany

More information about this series at <http://www.springer.com/series/1244>

Jinyan Li · Xue Li
Shuliang Wang · Jianxin Li
Quan Z. Sheng (Eds.)

Advanced Data Mining and Applications

12th International Conference, ADMA 2016
Gold Coast, QLD, Australia, December 12–15, 2016
Proceedings

Editors

Jinyan Li
University of Technology Sydney
Ultimo, NSW
Australia

Xue Li
University of Queensland
Brisbane, QLD
Australia

Shuliang Wang
Beijing Institute of Technology
Beijing
China

Jianxin Li
University of Western Australia
Crawley, WA
Australia

Quan Z. Sheng
University of Adelaide
Adelaide, SA
Australia

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-49585-9 ISBN 978-3-319-49586-6 (eBook)
DOI 10.1007/978-3-319-49586-6

Library of Congress Control Number: 2016957378

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing AG 2016

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, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The year 2016 marked the 12th anniversary of the International Conference on Advanced Data Mining and Applications (ADMA). This conference series has been organized and held in China since 2005. ADMA 2016 was the first time the conference was held outside China, taking place on Australia's Gold Coast during December 12–15, 2016. ADMA aims at bringing together both young researchers and senior experts from the world, and also at providing a leading international forum for the dissemination of original data mining findings and practical data mining experience. Over the years, ADMA has grown to become a flagship conference in the field of data mining and applications.

This volume comprises 57 thoroughly reviewed and carefully revised full-length papers that were all presented at the conference. The conference Program Committee received 105 high-quality manuscripts. Each of them was assigned to at least three Program Committee members for comments. All papers were rigorously reviewed and had at least three reviews. Some borderline papers were even reviewed by five members of the committee if the first three reviews were not adequate for making a rejection or acceptance decision. Several research teams showed strong interest in ADMA 2016, and submitted multiple high-quality papers of different topics. It was difficult for the committee chairs to exclude some of these papers so as to reserve some space for other teams to present their latest research achievements.

Although many papers were worthy of publication, only 18 spotlight research papers (acceptance rate of 17%, 15 pages each) and 39 regular research papers (13 pages each) could be accepted for presentation at the conference and publication in this volume. The selected papers covered a wide variety of important topics in the area of data mining, including parallel and distributed data mining algorithms, mining on data streams, graph mining, spatial data mining, multimedia data mining, Web mining, the Internet of Things, health informatics, and biomedical data mining. The conference program of ADMA 2016 was also complemented by four outstanding keynotes given by Yufei Tao, Pablo Moscato, Scott Johnston, and Gao Cong, as well as several research demonstrations. We would like to particularly thank the keynote speakers for presenting algorithms and ideas to solve data mining problems in the frontier areas of the field.

We thank the Program Committee members for their time and invaluable comments. The schedule of the reviewing process was extremely tight. The members' tremendous efforts to complete the review reports before the deadline are greatly appreciated. We read all of the review reports and found some of them to be truly excellent, as good as what is usually found in a survey—critical, digressive, and detailed. These comments were very helpful for us in selecting the papers. Thank you all and may the papers collected in this volume inspire the readers and open new windows of research.

We would like to express our gratitude to all individuals, institutions, and sponsors that supported ADMA 2016. This high-quality program would not have been possible without the expertise and dedication of our Program Committee members. We are

grateful for the guidance of the general chairs (Michael Sheng and Osmar Zaiane), the tireless efforts of the demonstration chairs (Ji Zhang and Mingkui Tan), the publicity chair (Xiujuan Xu), the publication chair (Jianxin Li), the award chair (Zhifeng Bao), the Web chair (Wenjie Ruan), the special issue chair (Yongrui Qin), and the local organization chair (Weitong Chen). We would also like to acknowledge the support of the members of the conference Steering Committee. All of them helped make ADMA 2016 a success. Finally, we would like to thank all researchers, practitioners, and students who contributed with their work and participated in the conference. We hope that you find the papers in the proceedings interesting and stimulating.

December 2016

Jinyan Li
Xue Li
Shuliang Wang

Organization

General Co-chairs

Michael Sheng	University of Adelaide, Australia
Osmar Zaiane	University of Alberta, Canada

Program Co-chairs

Jinyan Li	University of Technology Sydney, Australia
Xue Li	University of Queensland, Australia
Shuliang Wang	Beijing Institute of Technology, China

Demo Co-chairs

Mingkui Tan	University of Adelaide, Australia
Ji Zhang	University of Southern Queensland, Australia

Proceedings Chair

Jianxin Li	University of Western Australia, Australia
------------	--

Awards Committee Chair

Zhifeng Bao	RMIT, Australia
-------------	-----------------

Publicity Chair

Xiujuan Xu	Dalian University of Technology, China
------------	--

Special Issue Chair

Yongrui Qin	University of Huddersfield, UK
-------------	--------------------------------

Sponsorship Co-chairs

Shichao Zhang	Guangxi Normal University, China
Xianxiu Si	China Academy of Telecom Research, China

Local Chair

Weitong Chen	University of Queensland, Australia
--------------	-------------------------------------

Web Chair

Wenjie Ruan University of Adelaide, Australia

Steering Committee

Jie Cao	Nanjing University of Finance and Economics, China
Xue Li	University of Queensland, Australia (Chair)
Michael Sheng	University of Adelaide, Australia
Jie Tang	Tsinghua University, China
Kyu-Young Whang	Korea Advanced Institute of Science and Technology, Korea
Min Yao	Zhejiang University, China
Osmar Zaiane	University of Alberta, Canada
Chengqi Zhang	University of Technology Sydney, Australia
Shichao Zhang	Guangxi Normal University, China
Shuliang Wang	Beijing Institute of Technology, China

Program Committee

Ayan Acharya	University of Texas at Austin, USA
Djamal Benslimane	Lyon 1 University, France
Jin Chen	Michigan State University, USA
Liang Chen	RMIT University, Australia
Chenwei Deng	Beijing Institute of Technology, China
Xiangjun Dong	Qilu University of Technology, China
Xiu Susie Fang	University of Adelaide, Australia
Stefano Ferilli	University of Bari, Italy
Philippe Fournier-Viger	Harbin Institute of Technology Shenzhen Graduate School, China
Xiaoying Gao	Victoria University of Wellington, New Zealand
Sung Ho Ha	Kyungpook National University, South Korea
Jing He	Victoria University, Australia
Qing He	Institute of Computing Technology, CAS, China
Wei Hu	Nanjing University, China
Xiaohua Tony Hu	Drexel University, USA
Guangyan Huang	Deakin University, Australia
Akihiro Inokuchi	Kwansei Gakuin University, Japan
Daisuke Kawahara	Kyoto University, Japan
Gang Li	Deakin University, Australia
Haiquan Li	University of Arizona, USA
Jianxin Li	University of Western Australia
Jinyan Li	University of Technology Sydney, Australia
Xue Li	University of Queensland, Australia
Zhixin Li	Guangxi Normal University, China
Bo Liu	Guangdong University of Technology, China

Guohua Liu	Yanshan University, China
Yubao Liu	Sun Yat-Sen University, China
Xudong Luo	Sun Yat-Sen University, China
Jiangang Ma	Victoria University, Australia
Marco Maggini	University of Siena, Italy
Toshiro Minami	Kyushu Institute of Information Sciences and Kyushu University Library, Japan
Yasuhiko Morimoto	Hiroshima University, Japan
Pablo Moscato	University of Newcastle, Australia
Shirui Pan	University of Technology Sydney, Australia
Tao Peng	University of Illinois at Urbana-Champaign, USA
Shweta Purawat	San Diego Supercomputer Center, UCSD, USA
Tieyun Qian	Wuhan University, China
Kai Qin	RMIT University, Australia
Yongrui Qin	University of Huddersfield, UK
Wenjie Ruan	University of Adelaide, Australia
Michele Ruta	Politecnico di Bari, Italy
Arun Kumar Sangaiah	VIT University, India
Dharmendra Sharma	University of Canberra, Australia
Ali Shemshadi	University of Adelaide, Australia
Michael Sheng	University of Adelaide, Australia
Guojie Song	Peking University, Beijing, China
Eiji Uchino	Yamaguchi University, Japan
Li Wan	Google, USA
Hongzhi Wang	Harbin Institute of Technology, China
Qinsi Wang	Carnegie Mellon University, USA
Shuliang Wang	Beijing Institute of Technology, China
Xianzhi Wang	University of Adelaide, Australia
Zhihui Wang	Fudan University, China
Jia Wu	University of Technology, Sydney, Australia
Zhiang Wu	Nanjing University of Finance and Economics, China
Feng Xia	Dalian University of Technology, China
Zhipeng Xie	Fudan University, China
Guandong Xu	University of Technology Sydney, Australia
Xiujuan Xu	Dalian University of Technology, China
Bing Xue	Victoria University of Wellington, New Zealand
Zijiang Yang	York University, UK
Lina Yao	University of New South Wales, Australia
Min Yao	Zhejiang University, China
Dayong Ye	University of Wollongong, Australia
Chao Yu	University of Wollongong, Australia
Fusheng Yu	Beijing Normal University, China
Qi Yu	Rochester Institute of Technology, USA
Osmar R. Zaiane	University of Alberta, Canada
Chunxia Zhang	Beijing Institute of Technology, China

Guang Lan Zhang	Boston University, USA
Mengjie Zhang	Victoria University of Wellington, New Zealand
Shichao Zhang	Guangxi Normal University, China
Wei Emma Zhang	University of Adelaide, Australia
Xin Zhao	University of Queensland, Australia
Nenggan Zheng	Zhejiang University, China
Yong Zheng	Illinois Institute of Technology, USA
Mingyang Zhong	University of Queensland, Australia
Hill Zhu	Florida Atlantic University, USA
Xiaofeng Zhu	Guangxi Normal University, China
Yi Zhuang	Zhejiang Gongshang University, China

Additional Reviewers

Al-Sahaf, Harith	Quang Huy, Duong
Bai, Xiaomei	Rahim, Azizur
Bekele, Teshome Megersa	Siers, Michael
Boyu, Zhang	Sun, Zequn
Chen, Qi	Vu, Huy Quan
Dam, Thu-Lan	Wan, Yao
Fan, Wei	Wang, Biying
Gao, Qian	Wang, Dongjing
Gong, Saisai	Wang, Hongtao
Gueniche, Ted	Wang, Qingxiang
Hanane, Amirat	Wang, Wei
Keskin, Derin	Wu, Cheng Wei
Li, Xin	Wu, Runze
Liu, Shaowu	Xu, Peipei
Lu, Guangquan	Yang, Shuiqiao
Nguyen, Bach	Yu, Shuo
Nguyen, Thuong	Zhang, Boyu
Peng, Jiajie	Zhang, Yongshan
Qian, Xinqi	Zhao, Qi

Contents

Spotlight Research Papers

Effective Monotone Knowledge Integration in Kernel Support Vector Machines	3
<i>Christopher Bartley, Wei Liu, and Mark Reynolds</i>	
Textual Cues for Online Depression in Community and Personal Settings . . .	19
<i>Thin Nguyen, Svetha Venkatesh, and Dinh Phung</i>	
Confidence-Weighted Bipartite Ranking.	35
<i>Majdi Khalid, Indrakshi Ray, and Hamidreza Chitsaz</i>	
Mining Distinguishing Customer Focus Sets for Online Shopping Decision Support	50
<i>Lu Liu, Lei Duan, Hao Yang, Jyrki Nummenmaa, Guozhu Dong, and Pan Qin</i>	
Community Detection in Networks with Less Significant Community Structure	65
<i>Ba-Dung Le, Hung Nguyen, and Hong Shen</i>	
Prediction-Based, Prioritized Market-Share Insight Extraction	81
<i>Renato Keshet, Alina Maor, and George Kour</i>	
Interrelationships of Service Orchestrations.	95
<i>Victor W. Chu, Raymond K. Wong, Fang Chen, and Chi-Hung Chi</i>	
Outlier Detection on Mixed-Type Data: An Energy-Based Approach.	111
<i>Kien Do, Truyen Tran, Dinh Phung, and Svetha Venkatesh</i>	
Low-Rank Feature Reduction and Sample Selection for Multi-output Regression	126
<i>Shichao Zhang, Lifeng Yang, Yonggang Li, Yan Luo, and Xiaofeng Zhu</i>	
Biologically Inspired Pattern Recognition for E-nose Sensors	142
<i>Sanad Al-Maskari, Wenping Guo, and Xiaoming Zhao</i>	
Addressing Class Imbalance and Cost Sensitivity in Software Defect Prediction by Combining Domain Costs and Balancing Costs.	156
<i>Michael J. Siers and Md Zahidul Islam</i>	

Unsupervised Hypergraph Feature Selection with Low-Rank and Self-Representation Constraints	172
<i>Wei He, Xiaofeng Zhu, Yonggang Li, Rongyao Hu, Yonghua Zhu, and Shichao Zhang</i>	
Improving Cytogenetic Search with GPUs Using Different String Matching Schemes	188
<i>Chantana Chantrapornchai and Chidchanok Choksuchat</i>	
CEIoT: A Framework for Interlinking Smart Things in the Internet of Things	203
<i>Ali Shemshadi, Quan Z. Sheng, Yongrui Qin, and Ali Alzubaidi</i>	
Adopting Hybrid Descriptors to Recognise Leaf Images for Automatic Plant Specie Identification	219
<i>Ali A. Al-kharaz, Xiaohui Tao, Ji Zhang, and Raid Lafta</i>	
Efficient Mining of Pan-Correlation Patterns from Time Course Data	234
<i>Qian Liu, Jinyan Li, Limsoon Wong, and Kotagiri Ramamohanarao</i>	
Recognizing Daily Living Activity Using Embedded Sensors in Smartphones: A Data-Driven Approach	250
<i>Wenjie Ruan, Leon Chea, Quan Z. Sheng, and Lina Yao</i>	
Dynamic Reverse Furthest Neighbor Querying Algorithm of Moving Objects	266
<i>Bohan Li, Chao Zhang, Weitong Chen, Yingbao Yang, Shaohong Feng, Qiqian Zhang, Weiwei Yuan, and Dongjing Li</i>	
Research Papers	
Relative Neighborhood Graphs Uncover the Dynamics of Social Media Engagement	283
<i>Natalie Jane de Vries, Ahmed Shamsul Arefin, Luke Mathieson, Benjamin Lucas, and Pablo Moscato</i>	
An Ensemble Approach for Better Truth Discovery	298
<i>Xiu Susie Fang, Quan Z. Sheng, and Xianzhi Wang</i>	
Single Classifier Selection for Ensemble Learning	312
<i>Guangtao Wang, Xiaomei Yang, and Xiaoyan Zhu</i>	
Community Detection in Dynamic Attributed Graphs	329
<i>Gonzalo A. Bello, Steve Harenberg, Abhishek Agrawal, and Nagiza F. Samatova</i>	

Secure Computation of Skyline Query in <i>MapReduce</i>	345
<i>Asif Zaman, Md. Anisuzzaman Siddique, Annisa, and Yasuhiko Morimoto</i>	
Recommending Features of Mobile Applications for Developer	361
<i>Hong Yu, Yahong Lian, Shuotao Yang, Linlin Tian, and Xiaowei Zhao</i>	
Adaptive Multi-objective Swarm Crossover Optimization for Imbalanced Data Classification	374
<i>Jinyan Li, Simon Fong, Meng Yuan, and Raymond K. Wong</i>	
Causality-Guided Feature Selection	391
<i>Mandar S. Chaudhary, Doel L. Gonzalez II, Gonzalo A. Bello, Michael P. Angus, Dhara Desai, Steve Harenberg, P. Murali Doraiswamy, Fredrick H. M. Semazzi, Vipin Kumar, and Nagiza F. Samatova, for the Alzheimer's Disease Neuroimaging Initiative</i>	
Temporal Interaction Biased Community Detection in Social Networks	406
<i>Noha Alduaiji, Jianxin Li, Amitava Datta, Xiaolu Lu, and Wei Liu</i>	
Extracting Key Challenges in Achieving Sobriety Through Shared Subspace Learning	420
<i>Haripriya Harikumar, Thin Nguyen, Santu Rana, Sunil Gupta, Ramachandra Kaimal, and Svetha Venkatesh</i>	
Unified Weighted Label Propagation Algorithm Using Connection Factor	434
<i>Xin Wang, Songlei Jian, Kai Lu, and Xiaoping Wang</i>	
MetricRec: Metric Learning for Cold-Start Recommendations.	445
<i>Furong Peng, Xuan Lu, Jianfeng Lu, Richard Yi-Da Xu, Cheng Luo, Chao Ma, and Jingyu Yang</i>	
Time Series Forecasting on Engineering Systems Using Recurrent Neural Networks	459
<i>Dongxu Shao, Tianyou Zhang, Kamal Mannar, and Yue Han</i>	
EDAHT: An Expertise Degree Analysis Model for Mass Comments in the E-Commerce System	472
<i>Jiang Zhong, You Xiong, Weili Guo, and Jingyi Xie</i>	
A Scalable Document-Based Architecture for Text Analysis	481
<i>Ciprian-Octavian Truică, Jérôme Darmont, and Julien Velcin</i>	
DAPPFC: Density-Based Affinity Propagation for Parameter Free Clustering	495
<i>Hanning Yuan, Shuliang Wang, Yang Yu, and Ming Zhong</i>	

Effective Traffic Flow Forecasting Using Taxi and Weather Data	507
<i>Xiujuan Xu, Benzhe Su, Xiaowei Zhao, Zhenzhen Xu, and Quan Z. Sheng</i>	
Understanding Behavioral Differences Between Short and Long-Term Drinking Abstainers from Social Media	520
<i>Haripriya Harikumar, Thin Nguyen, Sunil Gupta, Santu Rana, Ramachandra Kaimal, and Svetha Venkatesh</i>	
Discovering Trip Hot Routes Using Large Scale Taxi Trajectory Data	534
<i>Linjiang Zheng, Qisen Feng, Weining Liu, and Xin Zhao</i>	
Discovering Spatially Contiguous Clusters in Multivariate Geostatistical Data Through Spectral Clustering	547
<i>Francky Fouedjio</i>	
On Improving Random Forest for Hard-to-Classify Records	558
<i>Md Nasim Adnan and Md Zahidul Islam</i>	
Scholarly Output Graph: A Graphical Article-Level Metric Indicating the Impact of a Scholar's Publications	567
<i>Yu Liu, Dan Lin, Jing Li, and Shimin Shan</i>	
Distributed Lazy Association Classification Algorithm Based on Spark	580
<i>Xueming Li, Chaoyang Zhang, Guangwei Chen, Xiaoteng Sun, Qi Zhang, and Haomin Yang</i>	
Event Evolution Model Based on Random Walk Model with Hot Topic Extraction	591
<i>Chunzi Wu, Bin Wu, and Bai Wang</i>	
Knowledge-Guided Maximal Clique Enumeration	604
<i>Steve Harenberg, Ramona G. Seay, Gonzalo A. Bello, Rada Y. Chirkova, P. Murali Doraiswamy, and Nagiza F. Samatova</i>	
Got a Complaint?- Keep Calm and Tweet It!	619
<i>Nitish Mittal, Swati Agarwal, and Ashish Sureka</i>	
Query Classification by Leveraging Explicit Concept Information	636
<i>Fang Wang, Ze Yang, Zhoujun Li, and Jianshe Zhou</i>	
Stabilizing Linear Prediction Models Using Autoencoder	651
<i>Shivapratap Gopakumar, Truyen Tran, Dinh Phung, and Svetha Venkatesh</i>	
Mining Source Code Topics Through Topic Model and Words Embedding. . .	664
<i>Wei Emma Zhang, Quan Z. Sheng, Ermyas Abebe, M. Ali Babar, and Andi Zhou</i>	

IPC Multi-label Classification Based on the Field Functionality of Patent Documents	677
<i>Sora Lim and YongJin Kwon</i>	
Unsupervised Component-Wise EM Learning for Finite Mixtures of Skew t -distributions	692
<i>Sharon X. Lee and Geoffrey J. McLachlan</i>	
Supervised Feature Selection by Robust Sparse Reduced-Rank Regression . . .	700
<i>Rongyao Hu, Xiaofeng Zhu, Wei He, Jilian Zhang, and Shichao Zhang</i>	
PUEPro: A Computational Pipeline for Prediction of Urine Excretory Proteins	714
<i>Yan Wang, Wei Du, Yanchun Liang, Xin Chen, Chi Zhang, Wei Pang, and Ying Xu</i>	
Partitioning Clustering Based on Support Vector Ranking	726
<i>Qing Peng, Yan Wang, Ge Ou, Yuan Tian, Lan Huang, and Wei Pang</i>	
Global Recursive Based Node Importance Evaluation	738
<i>Lu Zhao, Li Xiong, and Shan Xue</i>	
Extreme User and Political Rumor Detection on Twitter	751
<i>Cheng Chang, Yihong Zhang, Claudia Szabo, and Quan Z. Sheng</i>	
Deriving Public Sector Workforce Insights: A Case Study Using Australian Public Sector Employment Profiles	764
<i>Shameek Ghosh, Yi Zheng, Thorsten Lammers, Ying Ying Chen, Carolyn Fitzmaurice, Scott Johnston, and Jinyan Li</i>	
Real-Time Stream Mining Electric Power Consumption Data Using Hoeffding Tree with Shadow Features	775
<i>Simon Fong, Meng Yuen, Raymond K. Wong, Wei Song, and Kyungeun Cho</i>	
Real-Time Investigation of Flight Delays Based on the Internet of Things Data	788
<i>Abdulwahab Aljubairy, Ali Shemshadi, and Quan Z. Sheng</i>	
Demo Papers	
IRS-HD: An Intelligent Personalized Recommender System for Heart Disease Patients in a Tele-Health Environment	803
<i>Raid Lafta, Ji Zhang, Xiaohui Tao, Yan Li, and Vincent S. Tseng</i>	

Sentiment Analysis for Depression Detection on Social Networks	807
<i>Xiaohui Tao, Xujuan Zhou, Ji Zhang, and Jianming Yong</i>	
Traffic Flow Visualization Using Taxi GPS Data	811
<i>Xiujuan Xu, Zhenzhen Xu, and Xiaowei Zhao</i>	
Erratum to: Advanced Data Mining and Applications	E1
<i>Jinyan Li, Xue Li, Shuliang Wang, Jianxin Li, and Quan Z. Sheng</i>	
Author Index	815