Lecture Notes in Artificial Intelligence 7121

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

Jie Tang Irwin King Ling Chen
Jianyong Wang (Eds.)

Advanced Data Mining and Applications

7th International Conference, ADMA 2011
Beijing, China, December 17-19, 2011
Proceedings, Part II

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Jie Tang
Jianyong Wang
Tsinghua University
Department of Computer Science and Technology
Beijing, 100084, China
E-mail: {jietang, jianyong}@tsinghua.edu.cn

Irwin King
The Chinese University of Hong Kong
Department of Computer Science and Engineering
Hong Kong, SAR, China
E-mail: king@cse.cuhk.edu.hk

Ling Chen
University of Technology
Faculty of Engineering and Information Technology
Sydney, NSW 2007, Australia
E-mail: ling.chen@uts.edu.au

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-25855-8 e-ISBN 978-3-642-25856-5
DOI 10.1007/978-3-642-25856-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2, H.3, H.4, H.2.8, J.1, F.1, I.4

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2011
This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.
The use of general descriptive names, registered names, trademarks, 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.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

Preface

The continuous growth of digital technologies leads to not only the availability of massive amounts of data, but also the emergence of new types of data with novel characteristics. It poses new challenges for the data-mining research community to develop sophisticated data-mining algorithms as well as successful data-mining applications. For the purpose of promoting the original research in advanced data mining and applications, bringing together the experts on data mining throughout the world, and providing a leading international forum to exchange research ideas and results in emergent data-mining problems, the 7th International Conference on Advanced Data Mining and Applications was held in Beijing, China, in 2011.

The conference received 191 paper submissions from 47 countries and areas. All papers were peer reviewed by at least three members of the Program Committee (PC) composed of international experts in data-mining fields, as well as one Vice PC Co-chair. The PC, together with our PC Co-chairs, worked very hard to select papers through a rigorous review process and extensive discussion, and finally composed a diverse and exciting program including 35 full papers and 29 short papers. The ADMA 2011 program was highlighted by three keynote speeches from outstanding researchers in advanced data-mining and application areas: Philip S. Yu (University of Illinois Chicago), Wolfgang Nejdl (L3S Research Center), and Stefan Decker(National University of Ireland).

Without the support of several funding agencies and organizations, the successful organization of the ADMA 2011 would not be possible. These include sponsorships from: IBM Research, China Samsung Telecom R&D Center, and Tsinghua University. We would also like to express our gratitude to the General Co-chairs for all their precious advice and the Organizing Committee for their dedicated organizing efforts. Last but not least, we sincerely thank all the authors, presenters and attendees who jointly contributed to the success of ADMA 2011!

December 2011

Jie Tang
Irwin King
Ling Chen
Jianyong Wang

Organization

ADMA 2011 was organized by Tsinghua University, China, and the School of Information Technology and Electrical Engineering, University of Queensland, Australia.

Organizing Committee

Steering Committee Chair

Xue Li	University of Queensland, Australia

General Co-chairs

Deyi Li	Chinese Academy of Engineering, China
Bing Liu	University of Illinois at Chicago, USA
Charu C. Aggarwal	IBM T.J. Watson Research Center, USA

Program Co-chairs

Jie Tang	Tsinghua University, China
Jianyong Wang	Tsinghua University, China
Irwin King	The Chinese University of Hong Kong, China

Local Co-chair

Jun He	Renmin University of China, China

Regional Organization Co-chairs

Ruoming Jin	Kent State University, USA
Ee-Peng Lim	Singapore Management University, Singapore
Marie-Francine Moens	Katholieke Universiteit Leuven, Belgium
Jimeng Sun	IBM T.J. Watson Research Center, USA
Hwanjo Yu	Pohang University of Science and Technology, Korea
Xingquan Zhu	University of Technology Sydney, Australia

Proceedings Co-chairs

Ling Chen	University of Technology Sydney, Australia
Guoliang Li	Tsinghua University, China

Sponsor Co-chairs

Minlie Huang	Tsinghua University, China
Li Zhang	University of Wisconsin-Madison, USA

Finance chair

Peng Cui — Tsinghua University, China

Publicity Co-chairs

Juanzi Li — Tsinghua University, China
Zhichun Wang — Tsinghua University, China
Jibing Gong — Chinese Academy of Sciences, China

Industrial Track Chair

Keke Cai — IBM Research, China

Registration Chair

Xiaonan Liu — Tsinghua University, China

Special Track Co-chairs

Hongyan Liu — Tsinghua University, China
Zhong Su — IBM Research, China

Web Master

Bo Gao — Tsinghua University, China

Program Committee

Vice PC Chairs

Ling Chen, Australia
Hong Chen, Hong Kong
Xiaoyong Du, China
Zhan-huai Li, China
Marie-Francine Moens, Belgium
Yizhou Sun, USA
Hao Wang, USA
Aoying Zhou, China
Wei Chen, China
Bin Cui, China
Ruoming Jin, USA
Xiaofeng Meng, China
Kyuseok Shim, Korea
Wei Wang, China
Ying Zhao, China

PC Members

Aijun An, Canada
Akihiro Inokuchi, Japan
Ali Daud, China
Andrzej Skowron, Poland
Atsuyoshi Nakamura, Japan
Bruno Cremilleux, France
Cheqing Jin, China
Chiranjib Bhattacharyya, India
Chun-Hung Li, Hong Kong
Cindy Lin, USA
Aixin Sun, Singapore
Alfredo Cuzzocrea, Italy
Amanda Clare, UK
Annalisa Appice, Italy
Brijesh Verma, Australia
Chenhao Tan, USA
Chi Wang, USA
Chotirat Ratanamahatana, Thailand
Chun-Nan Hsu, USA
Daisuke Ikeda, Japan

Daisuke Kawahara, Japan
Daoqiang Zhang, China
David Taniar, Australia
Di Wu, China
Dianhui Wang, Australia
Du Zhang, USA
Faizah Shaari, Malaysia
Fusheng Yu, China
Gang Li, Australia
Guohe Li, China
Guoqiong Liao, China
Hanghang Tong, USA
Harry Zhang, Canada
Hongan Wang, China
Hongzhi Wang, China
Huan Huo, China
Huidong Jin, Australia
James Bailey, Australia
Jan Rauch, Czech Republic
Jiakui Zhao, China
Jianhui Chen, USA
Jibing Gong, China
Jinbao Li, China
Jing Liu, China
Jizhou Luo, China
Keke Cai, China
Kritsada Sriphaew, Japan
Lian Yu, China
Licheng Jiao, China
Lisa Hellerstein, USA
Mao Ye, China
Mario Linares-Vásquez, Colombia
Masashi Sugiyama, Japan
Mengjie Zhang, New Zealand
Michael Madden, Ireland
Michele Berlingerio, Italy
Min Yao, China
Ming Li, China
Ming-Syan Chen, Taiwan
Nicolas Spyratos, France
Odysseas Papapetro, Greece
Panagiotis Karras, Singapore
Philippe Fournier-Viger, Taiwan
Qinbao Song, China
Qingshan Liu, China
Danzhou Liu, USA
Dao-Qing Dai, China
Dexi Liu, China
Diane Cook, USA
Donghui Zhang, USA
Eduardo Hruschka, Brazil
Feiping Nie, USA
Gaël Dias, Portugal
George Karypis, USA
Guojie Song, China
Guoyin Wang, China
Hanzi Wang, China
Hassan Abolhassani, Iran
Hongyan Li, China
Hua Lu, Denmark
Hui Xiong, USA
Hung Son Nguyen, Poland
James Kwok, Hong Kong
Jason Wang, USA
Jian Yin, China
Jian-Min Han, China
Jimmy Huang, Canada
Jing Gao, USA
Jinghai Rao, China
K. Selcuk Candan, USA
Kitsana Waiyamai, Thailand
Li Li, China
Liang Sun, USA
Lidan Shou, China
Manish Gupta, USA
Marco Maggini, Italy
Martine De Cock, Belgium
Masayuki Numao, Japan
Michael Bain, Australia
Michalis Vazirgiannis, Greece
Michele Sebag, France
Ming Ji, China
Mingchun Wang, Taiwan
Nicola Di Mauro, Italy
Ninghui Li, USA
Pablo Castro, Argentina
Patrick Gallinari, France
Qi Wang, China
Qing He, China
Ran Wolff, Israel

Ravi Kumar, USA
Sadok Ben Yahia, Tunisia
Sang-Hyuk Lee, Korea
Sanparith Marukatat, Thailand
Sheng Zhong, USA
Shengtao Sun, China
Shu-Ching Chen, USA
Shuliang Wang, China
Songhua Xu, USA
Stefan Skudlarek, Japan
Sung Ho Ha, Republic of Korea
Takehisa Yairi, Japan
Tao Qin, China,
Thepchai Supnithi, Thailand
Tim Weninger, USA
Tomonari Masada, Japan
Tru Cao, Vietnam
Tu-Anh Nguyen-Hoang, Vietnam
Wai Lam, Hong Kong
Weizhu Chen, China
Wenyang Bai, China
Wlodek Zadrozny, USA
Xiangliang Zhang, Saudi Arabia
Xiaohui Liu, UK
Xin Jin, USA
Xintao Wu, USA
Xue Li, Australia
Yang Xiang, USA
Yasuhiko Morimoto, Japan
Yi Chen, USA
Yihua Wu, USA
Yi-Ping Chen, Australia
Yongli Wang, China
Yubao Liu, China
Yuhua Li, China
Zhihai Wang, China
Zhipeng Xie, China
Zhongzhi Shi, China
Zi Yang, China
Zili Zhang, Australia
Rui Camacho, Portugal
Sai Wu, Singapore
Sanjay Jain, Singapore
Shen Jialie, Singapore
Shengfei Shi, China
Shichao Zhang, China
Shuigeng Zhou, China
Songcan Chen, China
Srikanta Tirthapura, USA
Stefano Ferilli, Italy
Tadashi Nomoto, Japan
Tao Li, USA
Tetsuya Yoshida, Japan
Tieyun Qian, China
Tomoharu Iwata, Japan
Toshiro Minami, Japan
Tsuyoshi Murata, Japan
Wagner Meira Jr., Brazil
Wei Liu, Australia
Wen-Chih Peng, Taiwan
Wilfred Ng, Hong Kong
Wynne Hsu, Singapore
Xiaohua Hu, USA
Xide Lin, USA
Xingquan Zhu, Australia
Xiuli Ma, China
Xuelong Li, China
Yang-Sae Moon, Republic of Korea
Yasuhito Asano, Japan
Yifeng Zeng, Denmark
Ying Zhang, Australia
Yonghong Peng, UK
Yu Jian, China
Yueguo Chen, China
Zhenying He, China
Zhihong Deng, China
Zhongfei Zhang, USA
Zi Huang, Australia
Zijiang Yang, Canada

Table of Contents – Part II

Generating Syntactic Tree Templates for Feature-Based Opinion Mining 1
Liang Wu, Yuanchun Zhou, Fei Tan, Fenglei Yang, and Jianhui Li

Handling Concept Drift via Ensemble and Class Distribution Estimation Technique 13
Nachai Limsetto and Kitsana Waiyamai

HUE-Stream: Evolution-Based Clustering Technique for Heterogeneous Data Streams with Uncertainty 27
Wicha Meesuksabai, Thanapat Kangkachit, and Kitsana Waiyamai

Hybrid Artificial Immune Algorithm and CMAC Neural Network Classifier for Supporting Business and Medical Decision Making 41
Jui-Yu Wu

Improving Suffix Tree Clustering with New Ranking and Similarity Measures 55
Phiradit Worawitphinyo, Xiaoying Gao, and Shahida Jabeen

Individual Doctor Recommendation Model on Medical Social Network 69
Jibing Gong and Shengtao Sun

Influence Maximizing and Local Influenced Community Detection Based on Multiple Spread Model 82
Qiuling Yan, Shaosong Guo, and Dongqing Yang

Interactive Predicate Suggestion for Keyword Search on RDF Graphs . . . 96
Mengxia Jiang, Yueguo Chen, Jinchuan Chen, and Xiaoyong Du

Intrinsic Dimension Induced Similarity Measure for Clustering 110
Yu Xiao, Jian Yu, and Shu Gong

Learning to Make Social Recommendations: A Model-Based Approach 124
Xiongcai Cai, Michael Bain, Alfred Krzywicki, Wayne Wobcke, Yang Sok Kim, Paul Compton, and Ashesh Mahidadia

Microgroup Mining on TSina via Network Structure and User Attribute 138
Xiaobing Xiong, Xiang Niu, Gang Zhou, Ke Xu, and Yongzhong Huang

Mining Good Sliding Window for Positive Pathogens Prediction in Pathogenic Spectrum Analysis 152
Lei Duan, Changjie Tang, Chi Gou, Min Jiang, and Jie Zuo

Mining Patterns from Longitudinal Studies 166
Aída Jiménez, Fernando Berzal, and Juan-Carlos Cubero

Mining Top-K Sequential Rules 180
Philippe Fournier-Viger and Vincent S. Tseng

Mining Uncertain Data Streams Using Clustering Feature Decision Trees 195
Wenhua Xu, Zheng Qin, Hao Hu, and Nan Zhao

Multi-view Laplacian Support Vector Machines 209
Shiliang Sun

New Developments of Determinacy Analysis 223
Rein Kuusik and Grete Lind

On Mining Anomalous Patterns in Road Traffic Streams 237
Linsey Xiaolin Pang, Sanjay Chawla, Wei Liu, and Yu Zheng

Ontology Guided Data Linkage Framework for Discovering Meaningful Data Facts 252
Mohammed Gollapalli, Xue Li, Ian Wood, and Guido Governatori

Predicting New User's Behavior in Online Dating Systems 266
Tingting Wang, Hongyan Liu, Jun He, Xuan Jiang, and Xiaoyong Du

Sequential Pattern Mining from Stream Data 278
Adam Koper and Hung Son Nguyen

Social Influence Modeling on Smartphone Usage 292
Masaji Katagiri and Minoru Etoh

Social Network Inference of Smartphone Users Based on Information Diffusion Models 304
Tomonobu Ozaki and Minoru Etoh

Support Vector Regression with A Priori Knowledge Used in Order Execution Strategies Based on VWAP 318
Marcin Orchel

Terrorist Organization Behavior Prediction Algorithm Based on Context Subspace 332
Anrong Xue, Wei Wang, and Mingcai Zhang

Topic Discovery and Topic-Driven Clustering for Audit Method Datasets 346
Ying Zhao, Wanyu Fu, and Shaobin Huang

Transportation Modes Identification from Mobile Phone Data Using Probabilistic Models 359
Dafeng Xu, Guojie Song, Peng Gao, Rongzeng Cao, Xinwei Nie, and Kunqing Xie

User Graph Regularized Pairwise Matrix Factorization for Item Recommendation 372
Liang Du, Xuan Li, and Yi-Dong Shen

Using Predicate-Argument Structures for Context-Dependent Opinion Retrieval 386
Sylvester Olubolu Orimaye, Saadat M. Alhashmi, and Siew Eu-Gene

XML Document Clustering Using Structure-Preserving Flat Representation of XML Content and Structure 403
Fedja Hadzic, Michael Hecker, and Andrea Tagarelli

Author Index 417

Table of Contents – Part I

Retrieval in CBR Using a Combination of Similarity and Association Knowledge 1
Yong-Bin Kang, Shonali Krishnaswamy, and Arkady Zaslavsky

A Clustering Approach Using Weighted Similarity Majority Margins . . . 15
Raymond Bisdorff, Patrick Meyer, and Alexandru-Liviu Olteanu

A False Negative Maximal Frequent Itemset Mining Algorithm over Stream 29
Haifeng Li and Ning Zhang

A Graph Enrichment Based Clustering over Vertically Partitioned Data 42
Khalid Benabdeslem, Brice Effantin, and Haytham Elghazel

A Method for Finding Groups of Related Herbs in Traditional Chinese Medicine 55
Lidong Wang, Yin Zhang, Baogang Wei, Jie Yuan, and Xia Ye

A New Hybrid Clustering Method for Reducing Very Large Spatio-temporal Dataset 69
Michael Whelan, Nhien-An Le-Khac, and M.-Tahar Kechadi

A Normal Distribution-Based Over-Sampling Approach to Imbalanced Data Classification 83
Huaxiang Zhang and Zhichao Wang

A Novel Genetic Algorithm for Overlapping Community Detection 97
Yanan Cai, Chuan Shi, Yuxiao Dong, Qing Ke, and Bin Wu

A Probabilistic Topic Model with Social Tags for Query Reformulation in Informational Search 109
Yuqing Mao, Haifeng Shen, and Chengzheng Sun

A QoS-Aware Web Services Selection Model Using AND/OR Graph 124
Hong Yu and Man Liu

A Tweet-Centric Approach for Topic-Specific Author Ranking in Micro-Blog 138
Shoubin Kong and Ling Feng

An Algorithm for Sample and Data Dimensionality Reduction Using Fast Simulated Annealing 152
Szymon Łukasik and Piotr Kulczycki

An Investigation of Recursive Auto-associative Memory in Sentiment Detection 162
Saeed Danesh, Wei Liu, Tim French, and Mark Reynolds

APPECT: An Approximate Backbone-Based Clustering Algorithm for Tags 175
Yu Zong, Guandong Xu, Ping Jin, Yanchun Zhang, EnHong Chen, and Rong Pan

Bi-clustering Gene Expression Data Using Co-similarity 190
Syed Fawad Hussain

CCE: A Chinese Concept Encyclopedia Incorporating the Expert-Edited Chinese Concept Dictionary with Online Cyclopedias 201
Jiazhen Nian, Shan Jiang, Congrui Huang, and Yan Zhang

Cluster Ensembles via Weighted Graph Regularized Nonnegative Matrix Factorization 215
Liang Du, Xuan Li, and Yi-Dong Shen

Continuously Identifying Representatives Out of Massive Streams 229
Qiong Li, Xiuli Ma, Shiwei Tang, and Shuiyuan Xie

Cost-Sensitive Decision Tree for Uncertain Data 243
Mingjian Liu, Yang Zhang, Xing Zhang, and Yong Wang

Direct Marketing with Fewer Mistakes 256
Eileen A. Ni and Charles X. Ling

Discovering Collective Viewpoints on Micro-blogging Events Based on Community and Temporal Aspects 270
Bin Zhao, Zhao Zhang, Yanhui Gu, Xueqing Gong, Weining Qian, and Aoying Zhou

Discriminatory Confidence Analysis in Pattern Mining 285
Russel Pears, Yun Sing Koh, and Gillian Dobbie

Dominance-Based Soft Set Approach in Decision-Making Analysis 299
Awang Mohd Isa, Ahmad Nazari Mohd Rose, and Mustafa Mat Deris

Efficient Computation of Measurements of Correlated Patterns in Uncertain Data 311
Lisi Chen, Shengfei Shi, and Jing Lv

Efficient Subject-Oriented Evaluating and Mining Methods for Data with Schema Uncertainty 325
Yue Wang, Changjie Tang, Tengjiao Wang, Dongqing Yang, and Jun Zhu

An Empirical Evaluation of Bagging with Different Algorithms on Imbalanced Data 339
Guohua Liang and Chengqi Zhang

Exploiting Concept Clumping for Efficient Incremental News Article Categorization 353
Alfred Krzywicki and Wayne Wobcke

Extracting Rocks from Mars Images with Data Fields 367
Shuliang Wang and Yashen Chen

Finding a Wise Group of Experts in Social Networks 381
Hongzhi Yin, Bin Cui, and Yuxin Huang

Fully Utilize Feedbacks: Language Model Based Relevance Feedback in Information Retrieval 395
Sheng-Long Lv, Zhi-Hong Deng, Hang Yu, Ning Gao, and Jia-Jian Jiang

FXProj – A Fuzzy XML Documents Projected Clustering Based on Structure and Content 406
Tengfei Ji, Xiaoyuan Bao, and Dongqing Yang

Author Index 421