

# Lecture Notes in Artificial Intelligence 7120

## 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  
Jianying Wang (Eds.)

# Advanced Data Mining and Applications

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

## 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  
ISBN 978-3-642-25852-7  
DOI 10.1007/978-3-642-25853-4  
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349  
e-ISBN 978-3-642-25853-4

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

### **PC Members**

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

# Table of Contents – Part I

Retrieval in CBR Using a Combination of Similarity and Association Knowledge . . . . .	1
<i>Yong-Bin Kang, Shonali Krishnaswamy, and Arkady Zaslavsky</i>	
A Clustering Approach Using Weighted Similarity Majority Margins . . .	15
<i>Raymond Bisdorff, Patrick Meyer, and Alexandru-Liviu Olteanu</i>	
A False Negative Maximal Frequent Itemset Mining Algorithm over Stream . . . . .	29
<i>Haifeng Li and Ning Zhang</i>	
A Graph Enrichment Based Clustering over Vertically Partitioned Data . . . . .	42
<i>Khalid Benabdeslem, Brice Effantin, and Haytham Elghazel</i>	
A Method for Finding Groups of Related Herbs in Traditional Chinese Medicine . . . . .	55
<i>Lidong Wang, Yin Zhang, Baogang Wei, Jie Yuan, and Xia Ye</i>	
A New Hybrid Clustering Method for Reducing Very Large Spatio-temporal Dataset . . . . .	69
<i>Michael Whelan, Nhien-An Le-Khac, and M.-Tahar Kechadi</i>	
A Normal Distribution-Based Over-Sampling Approach to Imbalanced Data Classification . . . . .	83
<i>Huaxiang Zhang and Zhichao Wang</i>	
A Novel Genetic Algorithm for Overlapping Community Detection . . . .	97
<i>Yanan Cai, Chuan Shi, Yuxiao Dong, Qing Ke, and Bin Wu</i>	
A Probabilistic Topic Model with Social Tags for Query Reformulation in Informational Search . . . . .	109
<i>Yuqing Mao, Haifeng Shen, and Chengzheng Sun</i>	
A QoS-Aware Web Services Selection Model Using AND/OR Graph . . . .	124
<i>Hong Yu and Man Liu</i>	
A Tweet-Centric Approach for Topic-Specific Author Ranking in Micro-Blog . . . . .	138
<i>Shoubin Kong and Ling Feng</i>	
An Algorithm for Sample and Data Dimensionality Reduction Using Fast Simulated Annealing . . . . .	152
<i>Szymon Lukasik and Piotr Kulczycki</i>	

An Investigation of Recursive Auto-associative Memory in Sentiment Detection .....	162
<i>Saeed Danesh, Wei Liu, Tim French, and Mark Reynolds</i>	
APPECT: An Approximate Backbone-Based Clustering Algorithm for Tags .....	175
<i>Yu Zong, Guandong Xu, Ping Jin, Yanchun Zhang, EnHong Chen, and Rong Pan</i>	
Bi-clustering Gene Expression Data Using Co-similarity .....	190
<i>Syed Fawad Hussain</i>	
CCE: A Chinese Concept Encyclopedia Incorporating the Expert-Edited Chinese Concept Dictionary with Online Cyclopedias .....	201
<i>Jiazhen Nian, Shan Jiang, Congrui Huang, and Yan Zhang</i>	
Cluster Ensembles via Weighted Graph Regularized Nonnegative Matrix Factorization .....	215
<i>Liang Du, Xuan Li, and Yi-Dong Shen</i>	
Continuously Identifying Representatives Out of Massive Streams .....	229
<i>Qiong Li, Xiuli Ma, Shiwei Tang, and Shuiyuan Xie</i>	
Cost-Sensitive Decision Tree for Uncertain Data .....	243
<i>Mingjian Liu, Yang Zhang, Xing Zhang, and Yong Wang</i>	
Direct Marketing with Fewer Mistakes .....	256
<i>Eileen A. Ni and Charles X. Ling</i>	
Discovering Collective Viewpoints on Micro-blogging Events Based on Community and Temporal Aspects .....	270
<i>Bin Zhao, Zhao Zhang, Yanhui Gu, Xueqing Gong, Weining Qian, and Aoying Zhou</i>	
Discriminatory Confidence Analysis in Pattern Mining .....	285
<i>Russel Pears, Yun Sing Koh, and Gillian Dobbie</i>	
Dominance-Based Soft Set Approach in Decision-Making Analysis .....	299
<i>Awang Mohd Isa, Ahmad Nazari Mohd Rose, and Mustafa Mat Deris</i>	
Efficient Computation of Measurements of Correlated Patterns in Uncertain Data .....	311
<i>Lisi Chen, Shengfei Shi, and Jing Lv</i>	
Efficient Subject-Oriented Evaluating and Mining Methods for Data with Schema Uncertainty .....	325
<i>Yue Wang, Changjie Tang, Tengjiao Wang, Dongqing Yang, and Jun Zhu</i>	

An Empirical Evaluation of Bagging with Different Algorithms on Imbalanced Data . . . . .	339
<i>Guohua Liang and Chengqi Zhang</i>	
Exploiting Concept Clumping for Efficient Incremental News Article Categorization . . . . .	353
<i>Alfred Krzywicki and Wayne Wobcke</i>	
Extracting Rocks from Mars Images with Data Fields . . . . .	367
<i>Shuliang Wang and Yashen Chen</i>	
Finding a Wise Group of Experts in Social Networks . . . . .	381
<i>Hongzhi Yin, Bin Cui, and Yuxin Huang</i>	
Fully Utilize Feedbacks: Language Model Based Relevance Feedback in Information Retrieval . . . . .	395
<i>Sheng-Long Lv, Zhi-Hong Deng, Hang Yu, Ning Gao, and Jia-Jian Jiang</i>	
FXProj – A Fuzzy XML Documents Projected Clustering Based on Structure and Content . . . . .	406
<i>Tengfei Ji, Xiaoyuan Bao, and Dongqing Yang</i>	
<b>Author Index</b> . . . . .	421

## Table of Contents – Part II

Generating Syntactic Tree Templates for Feature-Based Opinion Mining .....	1
<i>Liang Wu, Yuanchun Zhou, Fei Tan, Fenglei Yang, and Jianhui Li</i>	
Handling Concept Drift via Ensemble and Class Distribution Estimation Technique .....	13
<i>Nachai Limsetto and Kitsana Waiyamai</i>	
HUE-Stream: Evolution-Based Clustering Technique for Heterogeneous Data Streams with Uncertainty .....	27
<i>Wicha Meesuksabai, Thanapat Kangkachit, and Kitsana Waiyamai</i>	
Hybrid Artificial Immune Algorithm and CMAC Neural Network Classifier for Supporting Business and Medical Decision Making .....	41
<i>Jui-Yu Wu</i>	
Improving Suffix Tree Clustering with New Ranking and Similarity Measures .....	55
<i>Phiradit Worawitphinyo, Xiaoying Gao, and Shahida Jabeen</i>	
Individual Doctor Recommendation Model on Medical Social Network .....	69
<i>Jibing Gong and Shengtao Sun</i>	
Influence Maximizing and Local Influenced Community Detection Based on Multiple Spread Model .....	82
<i>Qiuling Yan, Shaosong Guo, and Dongqing Yang</i>	
Interactive Predicate Suggestion for Keyword Search on RDF Graphs...	96
<i>Mengxia Jiang, Yueguo Chen, Jinchuan Chen, and Xiaoyong Du</i>	
Intrinsic Dimension Induced Similarity Measure for Clustering .....	110
<i>Yu Xiao, Jian Yu, and Shu Gong</i>	
Learning to Make Social Recommendations: A Model-Based Approach .....	124
<i>Xiongcai Cai, Michael Bain, Alfred Krzywicki, Wayne Wobcke, Yang Sok Kim, Paul Compton, and Ashesh Mahidadia</i>	
Microgroup Mining on TSina via Network Structure and User Attribute .....	138
<i>Xiaobing Xiong, Xiang Niu, Gang Zhou, Ke Xu, and Yongzhong Huang</i>	

Mining Good Sliding Window for Positive Pathogens Prediction in Pathogenic Spectrum Analysis . . . . .	152
<i>Lei Duan, Changjie Tang, Chi Gou, Min Jiang, and Jie Zuo</i>	
Mining Patterns from Longitudinal Studies . . . . .	166
<i>Aída Jiménez, Fernando Berzal, and Juan-Carlos Cubero</i>	
Mining Top-K Sequential Rules . . . . .	180
<i>Philippe Fournier-Viger and Vincent S. Tseng</i>	
Mining Uncertain Data Streams Using Clustering Feature Decision Trees . . . . .	195
<i>Wenhua Xu, Zheng Qin, Hao Hu, and Nan Zhao</i>	
Multi-view Laplacian Support Vector Machines . . . . .	209
<i>Shiliang Sun</i>	
New Developments of Determinacy Analysis . . . . .	223
<i>Rein Kuusik and Grete Lind</i>	
On Mining Anomalous Patterns in Road Traffic Streams . . . . .	237
<i>Linsey Xiaolin Pang, Sanjay Chawla, Wei Liu, and Yu Zheng</i>	
Ontology Guided Data Linkage Framework for Discovering Meaningful Data Facts . . . . .	252
<i>Mohammed Gollapalli, Xue Li, Ian Wood, and Guido Governatori</i>	
Predicting New User’s Behavior in Online Dating Systems . . . . .	266
<i>Tingting Wang, Hongyan Liu, Jun He, Xuan Jiang, and Xiaoyong Du</i>	
Sequential Pattern Mining from Stream Data . . . . .	278
<i>Adam Koper and Hung Son Nguyen</i>	
Social Influence Modeling on Smartphone Usage . . . . .	292
<i>Masaji Katagiri and Minoru Etoh</i>	
Social Network Inference of Smartphone Users Based on Information Diffusion Models . . . . .	304
<i>Tomonobu Ozaki and Minoru Etoh</i>	
Support Vector Regression with A Priori Knowledge Used in Order Execution Strategies Based on VWAP . . . . .	318
<i>Marcin Orchel</i>	
Terrorist Organization Behavior Prediction Algorithm Based on Context Subspace . . . . .	332
<i>Anrong Xue, Wei Wang, and Mingcai Zhang</i>	

Topic Discovery and Topic-Driven Clustering for Audit Method Datasets .....	346
<i>Ying Zhao, Wanyu Fu, and Shaobin Huang</i>	
Transportation Modes Identification from Mobile Phone Data Using Probabilistic Models .....	359
<i>Dafeng Xu, Guojie Song, Peng Gao, Rongzeng Cao, Xinwei Nie, and Kunqing Xie</i>	
User Graph Regularized Pairwise Matrix Factorization for Item Recommendation .....	372
<i>Liang Du, Xuan Li, and Yi-Dong Shen</i>	
Using Predicate-Argument Structures for Context-Dependent Opinion Retrieval .....	386
<i>Sylvester Olubolu Orimaye, Saadat M. Alhashmi, and Siew Eu-Gene</i>	
XML Document Clustering Using Structure-Preserving Flat Representation of XML Content and Structure .....	403
<i>Fedja Hadzic, Michael Hecker, and Andrea Tagarelli</i>	
<b>Author Index</b> .....	417