Lecture Notes in Artificial Intelligence 6441

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Longbing Cao
Yong Feng
Jiang Zhong (Eds.)

Advanced Data Mining and Applications

6th International Conference, ADMA 2010
Chongqing, China, November 19-21, 2010
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

Longbing Cao
University of Technology Sydney
Faculty of Engineering and Information Technology
Sydney, NSW 2007, Australia
E-mail: longbing.cao-1@uts.edu.au

Yong Feng
Chongqing University
College of Computer Science
Chongqing, 400030, China
E-mail: fengphd@msn.com

Jiang Zhong
Chongqing University
College of Computer Science
Chongqing, 400030, China
E-mail: zhongjiang@cqu.edu.cn

Library of Congress Control Number: 2010939048

CR Subject Classification (1998): I.2, H.3, H.4, H.2.8, J.1, H.5

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-642-17312-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-17312-7 Springer Berlin Heidelberg New York

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.

springer.com

© Springer-Verlag Berlin Heidelberg 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

With the ever-growing power of generating, transmitting, and collecting huge amounts of data, information overload is now an imminent problem to mankind. The overwhelming demand for information processing is not just about a better understanding of data, but also a better usage of data in a timely fashion. Data mining, or knowledge discovery from databases, is proposed to gain insight into aspects of data and to help people make informed, sensible, and better decisions. At present, growing attention has been paid to the study, development, and application of data mining. As a result there is an urgent need for sophisticated techniques and tools that can handle new fields of data mining, e.g., spatial data mining, biomedical data mining, and mining on high-speed and time-variant data streams. The knowledge of data mining should also be expanded to new applications.

The 6th International Conference on Advanced Data Mining and Applications (ADMA 2010) aimed to bring together the experts on data mining throughout the world. It provided a leading international forum for the dissemination of original research results in advanced data mining techniques, applications, algorithms, software and systems, and different applied disciplines. The conference attracted 361 online submissions from 34 different countries and areas. All full papers were peer reviewed by at least three members of the Program Committee composed of international experts in data mining fields. A total number of 118 papers were accepted for the conference. Amongst them, 63 papers were selected as regular papers and 55 papers were selected as short papers. The Program Committee worked very hard to select these papers through a rigorous review process and extensive discussion, and finally composed a diverse and exciting program for ADMA 2010. The ADMA 2010 program was highlighted by three keynote speeches from outstanding researchers in advanced data mining and application areas: Kotagiri Ramamohanarao, Chengqi Zhang, and Vladimir Brusic.

September 2010

Longbing Cao
Yong Feng
Jiang Zhong

Organization

ADMA 2010 was organized by Chongqing University, China, and the School of Information Technology and Electrical Engineering, the University of Queensland, Australia, sponsored by the National Natural Science Foundation of China, Chongqing Science and Technology Commission,Chongqing Academy of Science and Technology, and technically co-sponsored by IEEE Queensland Section.

Organizing Committee

Steering Committee Chair

Xue Li	University of Queensland, Australia

Keynote Speakers

Kotagiri Ramamohanarao	University of Melbourne, Australia
Chengqi Zhang	University of Technology Sydney, Australia
Vladimir Brusic	Dana Farber Cancer Institute, USA

General Co-chairs

Charles Ling	The University of Western Ontario, Canada
Shangbo Zhou	Chongqing University, China
Jie Xu	Leeds University, UK

Program Co-chairs

Jinyan Li	Nanyang Technological University, Singapore
Longbing Cao	University of Technology Sydney, Australia
Zhongyang Xiong	Chongqing University, China

Publicity Chair

Xueming Li	Chongqing University, China

Regional Organization Co-chairs

Jiang Zhong	Chongqing University, China
Yong Feng	Chongqing University, China

Finance Chair

Yong Feng	Chongqing University, China

China Registration Chair

Li Wan	Chongqing University, China

China Web Master

Quanji Qiu	Chongqing University, China
Xiaoran Lin	Chongqing University, China

China Secretariat

Min Zheng	Chongqing University, China
Yanyan Zou	Chongqing University, China
Haoyang Ren	Chongqing University, China
Junhui Wang	Chongqing University, China

Program Committee

Hua Li, Canada
Arlindo Oliveira, Portugal
Dragan Gamberger, Croatia
Andre Ponce Leao, Brazil
Andrew Kusiak, America
Wang Shuliang, China
Christophe Giraud-Carrier, USA
Daniel Neagu, UK
Liu Zhen, Japan
Daniel Sanchez, Spain
Dianhui Wang, Australia
Fernando Berzal, Spain
Gang Li, Australia
Jan Rauch, Czech Republic
Jean-Gabriel Ganascia, France
Joseph Roure, UK
Juho Rousu, USA
Junbin Gao, Australia
Paul Vitanyi, USA
Petr Berka, Czech Republic
Rui Camacho, Portugal
Wanquan Liu, Australia
Christophe Rigotti, India
Xiaochun Cheng, UK
Yonghong Peng, UK
Zhaoli Zhu, China
Peng Han, China
Cai Yueping, Japan
Yu Qiao, China
Guang Chen, China
Xinyang Ying, China
Guobin Zhou, China
Yun Li, China
Jun Zhao, China
Hong Tang, China
Hao Wang, China
Hong Yu, China
Li Li, China
Ling Ou, China
Zili Zhang, China
Xingang Zhang, China
Xiaofeng Liao, China
Kaigui Wu, China
Yufang Zhang, China
Hua Li, China
Xiaofan Yang, China
Jiang Zhong, China
Yong Feng, China
Ji Li, China
Li Wan, China
Chengliang Wang, China
Chunxiao Ye, China
Huiping Cao, USA
Yan Qi, USA
Yanchang Zhao, Australia
Sumon Shahriar, Australia

Senthil Kumar, India
Alfredo Cuzzocrea, Italy
Rong Xie, China
Ji Liu, China
Changze Wu, China
Yixiong Chen, China
Qi Xie, China
Pan He, China
BaoHua Qiang, China
Zhixing Li, China
Fenghua Tu, China
Xiaohong Zhang, China
Chuan Li, China

Table of Contents – Part II

III Data Mining Methodologies and Processes

Incremental Learning by Heterogeneous Bagging Ensemble 1
Qiang Li Zhao, Yan Huang Jiang, and Ming Xu

CPLDP: An Efficient Large Dataset Processing System Built on Cloud Platform 13
Zhiyong Zhong, Mark Li, Jin Chang, Le Zhou, Joshua Zhexue Huang, and Shengzhong Feng

A General Multi-relational Classification Approach Using Feature Generation and Selection 21
Miao Zou, Tengjiao Wang, Hongyan Li, and Dongqing Yang

A Unified Approach to the Extraction of Rules from Artificial Neural Networks and Support Vector Machines 34
João Guerreiro and Duarte Trigueiros

A Clustering-Based Data Reduction for Very Large Spatio-Temporal Datasets 43
Nhien-An Le-Khac, Martin Bue, Michael Whelan, and M-Tahar Kechadi

Change a Sequence into a Fuzzy Number 55
Diana Domańska and Marek Wojtylak

Multiple Kernel Learning Improved by MMD 63
Jiangtao Ren, Zhou Liang, and Shaofeng Hu

A Refinement Approach to Handling Model Misfit in Semi-supervised Learning 75
Hanjing Su, Ling Chen, Yunming Ye, Zhaocai Sun, and Qingyao Wu

Soft Set Approach for Selecting Decision Attribute in Data Clustering 87
Mohd Isa Awang, Ahmad Nazari Mohd Rose, Tutut Herawan, and Mustafa Mat Deris

Comparison of BEKK GARCH and DCC GARCH Models: An Empirical Study 99
Yiyu Huang, Wenjing Su, and Xiang Li

Adapt the mRMR Criterion for Unsupervised Feature Selection 111
Junling Xu

Evaluating the Distance between Two Uncertain Categorical Objects ... 122
Hongmei Chen, Lizhen Wang, Weiyi Liu, and Qing Xiao

Construction Cosine Radial Basic Function Neural Networks Based on Artificial Immune Networks ... 134
YongJin Zeng and JianDong Zhuang

Spatial Filter Selection with LASSO for EEG Classification ... 142
Wenting Tu and Shiliang Sun

Boolean Algebra and Compression Technique for Association Rule Mining ... 150
Somboon Anekritmongkol and M.L. Kulthon Kasamsan

Cluster Based Symbolic Representation and Feature Selection for Text Classification ... 158
B.S. Harish, D.S. Guru, S. Manjunath, and R. Dinesh

SimRate: Improve Collaborative Recommendation Based on Rating Graph for Sparsity ... 167
Li Yu, Zhaoxin Shu, and Xiaoping Yang

Logistic Regression for Transductive Transfer Learning from Multiple Sources ... 175
Yuhong Zhang, Xuegang Hu, and Yucheng Fang

Double Table Switch: An Efficient Partitioning Algorithm for Bottom-Up Computation of Data Cubes ... 183
Jinguo You, Lianying Jia, Jianhua Hu, Qingsong Huang, and Jianqing Xi

IV Data Mining Applications and Systems

Tag Recommendation Based on Bayesian Principle ... 191
Zhonghui Wang and Zhihong Deng

Comparison of Different Methods to Fuse Theos Images ... 202
Silong Zhang and Guojin He

Using Genetic K-Means Algorithm for PCA Regression Data in Customer Churn Prediction ... 210
Bingquan Huang, T. Satoh, Y. Huang, M.-T. Kechadi, and B. Buckley

Time-Constrained Test Selection for Regression Testing ... 221
Lian Yu, Lei Xu, and Wei-Tek Tsai

Chinese New Word Detection from Query Logs ... 233
Yan Zhang, Maosong Sun, and Yang Zhang

Exploiting Concept Clumping for Efficient Incremental E-Mail Categorization 244
Alfred Krzywicki and Wayne Wobcke

Topic-Based User Segmentation for Online Advertising with Latent Dirichlet Allocation 259
Songgao Tu and Chaojun Lu

Applying Multi-Objective Evolutionary Algorithms to QoS-Aware Web Service Composition 270
Li Li, Peng Cheng, Ling Ou, and Zili Zhang

Real-Time Hand Detection and Tracking Using LBP Features 282
Bin Xiao, Xiang-min Xu, and Qian-pei Mai

Modeling DNS Activities Based on Probabilistic Latent Semantic Analysis 290
Xuebiao Yuchi, Xiaodong Lee, Jian Jin, and Baoping Yan

A New Statistical Approach to DNS Traffic Anomaly Detection 302
Xuebiao Yuchi, Xin Wang, Xiaodong Lee, and Baoping Yan

Managing Power Conservation in Wireless Networks 314
Kongluan Lin, John Debenham, and Simeon Simoff

Using PCA to Predict Customer Churn in Telecommunication Dataset 326
T. Sato, B.Q. Huang, Y. Huang, M.-T. Kechadi, and B. Buckley

Hierarchical Classification with Dynamic-Threshold SVM Ensemble for Gene Function Prediction 336
Yiming Chen, Zhoujun Li, Xiaohua Hu, and Junwan Liu

Personalized Tag Recommendation Based on User Preference and Content 348
Zhaoxin Shu, Li Yu, and Xiaoping Yang

Predicting Defect Priority Based on Neural Networks 356
Lian Yu, Wei-Tek Tsai, Wei Zhao, and Fang Wu

Personalized Context-Aware QoS Prediction for Web Services Based on Collaborative Filtering 368
Qi Xie, Kaigui Wu, Jie Xu, Pan He, and Min Chen

Hybrid Semantic Analysis System – ATIS Data Evaluation 376
Ivan Habernal and Miloslav Konopík

Click Prediction for Product Search on C2C Web Sites 387
Xiangzhi Wang, Chunyang Liu, Guirong Xue, and Yong Yu

Finding Potential Research Collaborators in Four Degrees of Separation 399
Paweena Chaiwanarom, Ryutaro Ichise, and Chidchanok Lursinsap

Predicting Product Duration for Adaptive Advertisement 411
Zhongqi Guo, Yongqiang Wang, Gui-rong Xue, and Yong Yu

An Algorithm for Available Bandwidth Estimation of IPv6 Network 419
Quanjie Qiu, Zhiguo Li, and Zhongfu Wu

A Structure-Based XML Storage Method in YAFFS File System 427
Ji Liu, Shuyu Chen, and Haozhang Liu

A Multi-dimensional Trustworthy Behavior Monitoring Method Based on Discriminant Locality Preserving Projections 435
Guanghui Chang, Shuyu Chen, Huawei Lu, and Xiaoqin Zhang

NN-SA Based Dynamic Failure Detector for Services Composition in Distributed Environment 443
Changze Wu, Kaigui Wu, Li Feng, and Dong Tian

Two-Fold Spatiotemporal Regression Modeling in Wireless Sensor Networks 451
Hadi Shakibian and Nasrollah Moghadam Charkari

Generating Tags for Service Reviews 463
Suke Li, Jinmei Hao, and Zhong Chen

Developing Treatment Plan Support in Outpatient Health Care Delivery with Decision Trees Technique 475
Shahriyah Nyak Saad Ali, Ahmad Mahir Razali, Azuraliza Abu Bakar, and Nur Riza Suradi

Factor Analysis of E-business in Skill-Based Strategic Collaboration 483
Daijiang Chen and Juanjuan Chen

Increasing the Meaningful Use of Electronic Medical Records: A Localized Health Level 7 Clinical Document Architecture System 491
Jun Liang, Mei Fang Xu, Lan Juan Li, Sheng Li Yang, Bao Luo Li, De Ren Cheng, Ou Jin, Li Zhong Zhang, Long Wei Yang, and Jun Xiang Sun

Corpus-Based Analysis of the Co-occurrence of Chinese Antonym Pairs 500
Xingfu Wang, Zhongfu Wu, Yan Li, Qian Huang, and Jinglu Hui

Application of Decision-Tree Based on Prediction Model for Project Management 508
Xin-ying Tu and Tao Fu

Management Policies Analysis for Multi-core Shared Caches 514
Jianjun Du, Yixing Zhang, Zhongfu Wu, and Xinwen Wang

Multi-core Architecture Cache Performance Analysis and Optimization Based on Distributed Method 522
Kefei Cheng, Kewen Pan, Jun Feng, and Yong Bai

The Research on the User Experience of E-Commercial Website Based on User Subdivision 529
Wei Liu, Lijuan Lv, Daoli Huang, and Yan Zhang

An Ontology-Based Framework Model for Trustworthy Software Evolution 537
Ji Li, Chunmei Liu, and Zhiguo Li

Multi-level Log-Based Relevance Feedback Scheme for Image Retrieval 545
Huanchen Zhang, Weifeng Sun, Shichao Dong, Long Chen, and Chuang Lin

A Distributed Node Clustering Mechanism in P2P Networks 553
Mo Hai and Shuhang Guo

Exploratory Factor Analysis Approach for Understanding Consumer Behavior toward Using Chongqing City Card 561
Juanjuan Chen and Chengliang Wang

Author Index 569

Table of Contents – Part I

I Data Mining Foundations

Cost Sensitive Classification in Data Mining 1
Zhenxing Qin, Chengqi Zhang, Tao Wang, and Shichao Zhang

Web Users Access Paths Clustering Based on Possibilistic and Fuzzy Sets Theory 12
Hong Yu, Hu Luo, and Shuangshuang Chu

Discriminative Markov Logic Network Structure Learning Based on Propositionalization and χ^2-Test 24
Quang-Thang Dinh, Matthieu Exbrayat, and Christel Vrain

EWGen: Automatic Generation of Item Weights for Weighted Association Rule Mining 36
Russel Pears, Yun Sing Koh, and Gillian Dobbie

Best Clustering Configuration Metrics: Towards Multiagent Based Clustering 48
Santhana Chaimontree, Katie Atkinson, and Frans Coenen

On Probabilistic Models for Uncertain Sequential Pattern Mining 60
Muhammad Muzammal and Rajeev Raman

Cube Based Summaries of Large Association Rule Sets 73
Marie Ndiaye, Cheikh T. Diop, Arnaud Giacometti, Patrick Marcel, and Arnaud Soulet

A Perceptron-Like Linear Supervised Algorithm for Text Classification 86
Anestis Gkanogiannis and Theodore Kalamboukis

Research on Time Series Forecasting Model Based on Moore Automata 98
Yixiong Chen, Zhongfu Wu, Zhiguo Li, and Yixing Zhang

A Clustering Algorithm FCM-ACO for Supplier Base Management 106
Weining Liu and Lei Jiang

Nearest Neighbour Distance Matrix Classification 114
Mohd Shamrie Sainin and Rayner Alfred

Classification Inductive Rule Learning with Negated Features 125
Stephanie Chua, Frans Coenen, and Grant Malcolm

Fast Retrieval of Time Series Using a Multi-resolution Filter with Multiple Reduced Spaces 137
Muhammad Marwan Muhammad Fuad and Pierre-François Marteau

DHPTID-HYBRID Algorithm: A Hybrid Algorithm for Association Rule Mining 149
Shilpa Sonawani and Amrita Mishra

An Improved Rough Clustering Using Discernibility Based Initial Seed Computation 161
Djoko Budiyanto Setyohadi, Azuraliza Abu Bakar, and Zulaiha Ali Othman

Fixing the Threshold for Effective Detection of Near Duplicate Web Documents in Web Crawling 169
V.A. Narayana, P. Premchand, and A. Govardhan

Topic-Constrained Hierarchical Clustering for Document Datasets 181
Ying Zhao

Discretization of Time Series Dataset Using Relative Frequency and K-Nearest Neighbor Approach 193
Azuraliza Abu Bakar, Almahdi Mohammed Ahmed, and Abdul Razak Hamdan

MSDBSCAN: Multi-density Scale-Independent Clustering Algorithm Based on DBSCAN 202
Gholamreza Esfandani and Hassan Abolhassani

An Efficient Algorithm for Mining Erasable Itemsets 214
Zhihong Deng and Xiaoran Xu

Discord Region Based Analysis to Improve Data Utility of Privately Published Time Series 226
Shuai Jin, Yubao Liu, and Zhijie Li

Deep Web Sources Classifier Based on DSOM-EACO Clustering Model 238
Yong Feng, Xianyong Chen, and Zhen Chen

Kernel Based K-Medoids for Clustering Data with Uncertainty 246
Baoguo Yang and Yang Zhang

Frequent Pattern Mining Using Modified CP-Tree for Knowledge Discovery 254
R. Vishnu Priya, A. Vadivel, and R.S. Thakur

Spatial Neighborhood Clustering Based on Data Field 262
Meng Fang, Shuliang Wang, and Hong Jin

Surrounding Influenced K-Nearest Neighbors: A New Distance Based Classifier 270
I. Mendialdua, B. Sierra, E. Lazkano, I. Irigoien, and E. Jauregi

A Centroid k-Nearest Neighbor Method 278
Qingjiu Zhang and Shiliang Sun

Mining Spatial Association Rules with Multi-relational Approach 286
Min Qian, Li-Jie Pu, Rong Fu, and Ming Zhu

An Unsupervised Classification Method of Remote Sensing Images Based on Ant Colony Optimization Algorithm 294
Duo Wang and Bo Cheng

A Novel Clustering Algorithm Based on Gravity and Cluster Merging 302
Jiang Zhong, Longhai Liu, and Zhiguo Li

II Data Mining in Specific Areas

Evolution Analysis of a Mobile Social Network 310
Hao Wang and Alvin Chin

Distance Distribution and Average Shortest Path Length Estimation in Real-World Networks 322
Qi Ye, Bin Wu, and Bai Wang

Self-adaptive Change Detection in Streaming Data with Non-stationary Distribution 334
Xiangliang Zhang and Wei Wang

Anchor Points Seeking of Large Urban Crowd Based on the Mobile Billing Data 346
Wenhao Huang, Zhengbin Dong, Nan Zhao, Hao Tian, Guojie Song, Guanhua Chen, Yun Jiang, and Kunqing Xie

Frequent Pattern Trend Analysis in Social Networks 358
Puteri N.E. Nohuddin, Rob Christley, Frans Coenen, Yogesh Patel, Christian Setzkorn, and Shane Williams

Efficient Privacy-Preserving Data Mining in Malicious Model 370
Keita Emura, Atsuko Miyaji, and Mohammad Shahriar Rahman

Analyze the Wild Birds' Migration Tracks by MPI-Based Parallel Clustering Algorithm 383
HaiMing Zhang, YuanChun Zhou, JianHui Li, XueZhi Wang, and BaoPing Yan

Formal Concept Analysis Based Clustering for Blog Network Visualization 394
Jing Gao and Wei Lai

Finding Frequent Subgraphs in Longitudinal Social Network Data Using a Weighted Graph Mining Approach 405
Chuntao Jiang, Frans Coenen, and Michele Zito

Weigted-FP-Tree Based XML Query Pattern Mining 417
Mi Sug Gu, Jeong Hee Hwang, and Keun Ho Ryu

Privacy-Preserving Data Mining in Presence of Covert Adversaries 429
Atsuko Miyaji and Mohammad Shahriar Rahman

Multiple Level Views on the Adherent Cohesive Subgraphs in Massive Temporal Call Graphs 441
Qi Ye, Bin Wu, and Bai Wang

Combating Link Spam by Noisy Link Analysis 453
Yitong Wang, Xiaofei Chen, and Xiaojun Feng

High Dimensional Image Categorization 465
François Poulet and Nguyen-Khang Pham

Efficiently Mining Co-Location Rules on Interval Data 477
Lizhen Wang, Hongmei Chen, Lihong Zhao, and Lihua Zhou

Multiple Attribute Frequent Mining-Based for Dengue Outbreak 489
Zalizah Awang Long, Azuraliza Abu Bakar, Abdul Razak Hamdan, and Mazrura Sahani

A Top-Down Approach for Hierarchical Cluster Exploration by Visualization 497
Ke-Bing Zhang, Mehmet A. Orgun, Peter A. Busch, and Abhaya C. Nayak

Distributed Frequent Items Detection on Uncertain Data 509
Shuang Wang, Guoren Wang, and Jitong Chen

Mining Uncertain Sentences with Multiple Instance Learning 521
Feng Ji, Xipeng Qiu, and Xuanjing Huang

WeightLOFCC: A Heuristic Weight-Setting Strategy of LOF Applied to Outlier Detection in Time Series Data 529
Hongrui Xie, Yujiu Yang, and Wenhuang Liu

TGP: Mining Top-K Frequent Closed Graph Pattern without Minimum Support 537
Yuhua Li, Quan Lin, Ruixuan Li, and Dongsheng Duan

Research on Similarity Matching for Multiple Granularities Time-Series Data 549
Wenning Hao, Enlai Zhao, Hongjun Zhang, Gang Chen, and Dawei Jin

A Novel Algorithm for Hierarchical Community Structure Detection in Complex Networks 557
Chuan Shi, Jian Zhang, Liangliang Shi, Yanan Cai, and Bin Wu

Investigating Sequential Patterns of DNS Usage and Its Applications ... 565
Jun Wu, Xin Wang, Xiaodong Lee, and Baoping Yan

Key Issues and Theoretical Framework on Moving Objects Data Mining 577
Rong Xie and Xin Luo

An Improved KNN Based Outlier Detection Algorithm for Large Datasets 585
Qian Wang and Min Zheng

Some Developments of Determinacy Analysis 593
Rein Kuusik and Grete Lind

A New Computational Framework for Gene Expression Clustering 603
Shahreen Kasim, Safaai Deris, and Razib M. Othman

Forecasting Short-Term Trends of Stock Markets Based on Fuzzy Frequent Pattern Tree 611
Defu Zhang, Bo Wu, Xian Hua, and Yangbin Yang

Inspired Rule-Based User Identification 618
Peng Yang and Yan Zheng

Author Index 625