Lecture Notes in Artificial Intelligence 5139
Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Changjie Tang Charles X. Ling
Xiaofang Zhou Nick J. Cercone Xue Li (Eds.)

Advanced Data Mining and Applications

4th International Conference, ADMA 2008
Chengdu, China, October 8-10, 2008
Proceedings

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

Changjie Tang
Sichuan University, Computer School
Chengdu 610065, China
E-mail: tangchangjie@cs.scu.edu.cn

Charles X. Ling
The University of Western Ontario, Department of Computer Science
Ontario N6A 5B7, Canada
E-mail: cling@csd.uwo.ca

Xiaofang Zhou
Xue Li
The University of Queensland, School of Information Technology and
Electrical Engineering, Brisbane QLD 4072, Queensland, Australia
E-mail: {xueli; zxf}@itee.uq.edu.au

Nick J. Cercone
York University, Faculty of Science & Engineering
Toronto M3J 1P3 Ontario, Canada
E-mail: ncercone@yorku.ca

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2, H.2, H.2.8, H.3-4, K.4.4, J.3, I.4, J.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-88191-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-88191-9 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 is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2008
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12539390 06/3180 5 4 3 2 1 0

Preface

The Fourth International Conference on Advanced Data Mining and Applications (ADMA 2008) will be held in Chengdu, China, followed by the last three successful ADMA conferences (2005 in Wu Han, 2006 in Xi'an, and 2007 Harbin). Our major goal of ADMA is to bring together the experts on data mining in the world, and to provide a leading international forum for the dissemination of original research results in data mining, including applications, algorithms, software and systems, and different disciplines with potential applications of data mining. This goal has been partially achieved in a very short time despite the young age of the conference, thanks to the rigorous review process insisted upon, the outstanding list of internationally renowned keynote speakers and the excellent program each year. ADMA is ranked higher than, or very similar to, other data mining conferences (such as PAKDD, PKDD, and SDM) in early 2008 by an independent source: cs-conference-ranking.org.

This year we had the pleasure and honor to host illustrious keynote speakers. Our distinguished keynote speakers are Prof. Qiang Yang and Prof. Jiming Liu. Prof. Yang is a tenured Professor and postgraduate studies coordinator at Computer Science and Engineering Department of Hong Kong University of Science and Technology. He is also a member of AAAI, ACM, a senior member of the IEEE, and he is also an associate editor for the IEEE TKDE and IEEE Intelligent Systems, KAIS and WI Journals. Since 2002, he has published 27 journal papers and 53 conference papers including 8 top conferences such as AAAI, KDD, SIGIR, etc. Prof. Liu is Professor and Head of Computer Science Department at Hong Kong Baptist University. He was a tenured Professor and Director of School of Computer Science at University of Windsor, Canada. He has published over 200 research articles in refereed international journals and conferences, and a number of books. Prof. Liu has served academic and professional communities in various capacities, e.g., presently as Editor-in-Chief of Web Intelligence and Agent Systems, Associate Editor of IEEE Transactions on Knowledge and Data Engineering and Computational Intelligence, etc.

This year ADMA received totally 304 paper submissions from 21 different countries, making it, yet again, a truly international conference. A rigorous process of pre-screening and review involved 89 well-known international program committee members and 2 program co-chairs, in addition to numerous external reviewers. This screening process yielded the remarkable papers organized in these proceedings with 35 regular papers and 43 short papers, bearing a total acceptance rate of 25.6%.

Earthquakes on May 12th, 2008 changed the original schedule but never changed the authors' great support and the organizers' huge efforts to make ADMA succeed. During the hard days, we received numerous emails or calls asking and consoling about our situation. And the steering committee has given us enormous help and guidance. We have resumed work only days after the earthquake. With the help,

consideration and hard work of all organizers, authors, and conference attendees, ADMA 2008 will become another successful international conference in the data mining community.

July 2008

Changjie Tang
Charles X. Ling
Nick Cercone
Xiaofang Zhou
Xue Li

Organization

ADMA 2008 was organized by Sichuan University, China.

Steering Committee Chair

Xue Li	University of Queensland (UQ), Australia

General Co-chairs

Nick Cercone	York University, Canada
Xiaofang Zhou	Queensland University, Australia

Program Co-chairs

Changjie Tang	Sichuan University, China
Charles Ling	University of Western Ontario, Canada

Local Arrangements Co-chairs

Jiliu Zhou	Sichuan University, China
Chuan Li	Sichuan University, China

Publicity Co-chairs

Tao Li	Florida University, USA, UK
Xingshu Chen	Sichuan University, China

Finance Co-chairs

Guirong Xue	Shanghai Jiaotong University, China
Dou Shen	Microsoft Redmond AdLab

Registration Chair

Mei Hong	Sichuan University, China
Meiqi Liu	Sichuan University, China

Web Co-masters

Chunqiu Zeng	Sichuan University, China
Yue Zhang	Sichuan University, China

Steering Committee

Xue Li, University of Queensland, Australia
Email: xueli@itee.uq.edu.au
URL: http://www.itee.uq.edu.au/~xueli

Qiang Yang, Hong Kong University of Science and Technology, China
Email:qyang@cse.ust.hk
URL: http://www.cse.ust.hk/~qyang/

Whang, Kyu-Young, Korea Advanced Institute of Science and Technology, Korea
E-mail: kywhang@cs.kaist.ac.kr
URL: http://dblab.kaist.ac.kr/Prof/main_eng.html

Osmar R. Zaïane, University of Alberta, Canada
E-mail: zaiane@cs.ualberta.ca
URL: http://www.cs.ualberta.ca/~zaiane/

Chengqi Zhang, University of Technology, Sydney, Australia
E-mail: chengqi@it.uts.edu.au
URL: http://www-staff.it.uts.edu/~chengqi

Program Committee

Hassan Abolhassabni	Sharif University of Technology, Iran
Reda Alhajj	University of Calgary, Canada
James Bailey	University of Melbourne, Australia
Michael R. Berthold	University of Konstanz, Germany
Fernando Berzal	University of Granada, Spain
Jeremy Besson	Insa-Lyon, France
Francesco Bonchi	KDD Laboratory–ISTI CNR Pisa, Italy
Rui Camacho	University of Porto, Portugal
Nick Cercone	York University, Canada
Yu Chen	Sichuan University, China
Frans Coenen	University of Liverpool, UK
Alfredo Cuzzocrea	University of Calabria, Italy
Xiangjun Dong	Shandong Institute of Light Industry, China
Zhaoyang Dong	University of Queensland, Australia
Xiaoyong Du	Renmin University, China
Mohammad El-Hajj	University of Alberta, Canada

Ming Fan	Zhengzhou University, China
Yi Feng	Zhejiang University, China
Joao Gama	University of Porto, Portugal
Jean-Gabriel G. Ganascia	LIP6 - University Paris
Hong Gao	Harbin Institute of Technology, China
Junbin Gao	University of New England, Australia
Yu Ge	North East University, China
Peter Geczy	National Institute of Advanced Industrial Science and Technology (AIST), Japan
Christophe Giraud-Carrier	Brigham Young University
Vladimir Gorodetsky	Intelligent System Lab, Russian Academy of Science, Russia
Bing Guo	Sichuan University, China
Jimmy Huang	York University, Canada
Alfred Hofmann	Springer Verlag, Germany
Shengyi Jiang	GuangDong University of Foreign Studies
Yulan Ju	Chief Editor and Standing Deputy Editor-in-Chief of *Journal of Frontiers of Computer Science and Technology* (FCST)
Dimitrios Katsaros	Aristotle University, Greece
Mehmet Kaya	Firat University, Turkey
Adam Krzyzak	Concordia University, Montreal/Canada
Andrew Kusiak	University of Iowa, USA
Longin Jan Latecki	Temple University Philadelphia, USA
Gang Li	Deakin University, Australia
Yingshu Li	Georgia State University, USA
Zhanhuai Li	Northwest Polytechnical University, China
Chuan Li	Sichuan University, China
Xue Li	University of Queensland (UQ), Australia
Charles Ling	University of Western Ontario, Canada
Wanquan Liu	Curtin University of Technology, Australia
Jing Liu	Xidian University, China
Giuseppe Manco	National Research Council of Italy, Italy
Nasrullah Memon	Aalborg University, Denmark
Xiaofeng Meng	School of Information, Renmin University of China, China
Weiyi Meng	State University of New York at Binghamton, USA
Juggapong Natwichai	Chiang Mai University, Chiang Mai, Thailand
Daniel C. Neagu	University of Bradford, UK
Tansel Ozyer	TOBB University, Turkey
Deepak S. Padmanabhan	IBM India Research Lab
Jian Peng	Sichuan University, China
Yonghong Peng	University of Bradford, UK
Mithun Prasad	Rensselaer Polytechnic Institute, USA
Naren Ramakrishnan	Virginia Tech, USA
Zbigniew W. Ras	University of North Carolina, USA
Jan Rauch	University of Economics, Prague, Czech Republic

Raul Giraldez Rojo	Pablo de Olavide University, Spain
Ashkan Sami	Shiraz University, Iran
Giovanni Semeraro	University of Bari, Italy
Shengfei Shi	Harbin Institute of Technology, China
Carlos Soares	University of Porto, Portugal
Jaideep Srivastava	University of Minnesota, USA
Simsek Sule	University of Missouri-Rolla, USA
Kay Chen Tan	National University of Singapore, Singapore
Ah-Hwee Tan	Nanyang Technological University, Singapore
Changjie Tang	Sichuan University, China
Arthur Tay	National University of Singapore, Singapore
Luis Torgo	University of Porto, Portugal
Grigorios Tsoumakas	Aristotle University, Greece
Ricardo Vilalta	University of Houston, USA
Paul Vitanyi	CWI, The Netherlands
Wei Wang	Fudan University, China
Guoren Wang	NorthEast University, China
Shuliang Wang	Wuhan University, China
Desheng Dash Wu	University of Toronto, Canada
Zhipeng Xie	Fudan University, China
Qiang Yang	Hong Kong University of Science and Technology, Hong Kong
JingTao Yao	University of Regina, Canada
Jeffrey Xu Yu	Chinese University of Hong Kong, Hong Kong, China
Sarah Zelikovitz	College of Staten Island, NY, USA
Jianzhou Zhang	Sichuan University, China
Shichao Zhang	University of Technology, Sydney, Australia
Yang ZHANG	Northwest A&F University, China
Aoying Zhou	East China Normal University, China
Shuigeng Zhou	Fudan University, China
Xiaofang Zhou	University of Queensland (UQ), Australia

Sponsoring Institutions

National Science Foundation of China
WiseSoft Company Limited, Sichuan University

Table of Contents

Keynotes

An Introduction to Transfer Learning 1
Qiang Yang

Autonomy-Oriented Computing (AOC), Self-organized Computability, and Complex Data Mining 2
Jiming Liu

Regular Papers

Improving Angle Based Mappings 3
Frank Rehm and Frank Klawonn

Mining Natural Language Programming Directives with Class-Oriented Bayesian Networks 15
Manolis Maragoudakis, Nikolaos Cosmas, and Aristogiannis Garbis

Boosting over Groups and Its Application to Acronym-Expansion Extraction 27
Weijian Ni, Yalou Huang, Dong Li, and Yang Wang

A Genetic-Based Feature Construction Method for Data Summarisation 39
Rayner Alfred

Suicidal Risk Evaluation Using a Similarity-Based Classifier 51
S. Chattopadhyay, P. Ray, H.S. Chen, M.B. Lee, and H.C. Chiang

Gene Selection for Cancer Classification Using DCA 62
Hoai An Le Thi, Van Vinh Nguyen, and Samir Ouchani

FARS: A Multi-relational Feature and Relation Selection Approach for Efficient Classification 73
Bo Hu, Hongyan Liu, Jun He, and Xiaoyong Du

Enhancing Text Categorization Using Sentence Semantics 87
Shady Shehata, Fakhri Karray, and Mohamed Kamel

Mining Evolving Web Sessions and Clustering Dynamic Web Documents for Similarity-Aware Web Content Management 99
Jitian Xiao

Data Quality in Privacy Preservation for Associative Classification 111
Nattapon Harnsamut, Juggapong Natwichai, Xingzhi Sun, and Xue Li

Timeline Analysis of Web News Events 123
Jiangtao Qiu, Chuan Li, Shaojie Qiao, Taiyong Li, and Jun Zhu

Analysis of Alarm Sequences in a Chemical Plant 135
Savo Kordic, Peng Lam, Jitian Xiao, and Huaizhong Li

Speed Up SVM Algorithm for Massive Classification Tasks 147
Thanh-Nghi Do, Van-Hoa Nguyen, and François Poulet

Mining Supplemental Frequent Patterns 158
Yintian Liu, Yingming Liu, Tao Zeng, Kaikuo Xu, and Rong Tang

A Distributed Privacy-Preserving Association Rules Mining Scheme Using Frequent-Pattern Tree 170
Chunhua Su and Kouichi Sakurai

Dichotomy Method toward Interactive Testing-Based Fault Localization 182
Ji-Rong Sun, Zhi-Shu Li, and Jian-Cheng Ni

Maintaining the Maximum Normalized Mean and Applications in Data Stream Mining 194
Jan Peter Patist

Identification of Interface Residues Involved in Protein-Protein Interactions Using Naïve Bayes Classifier 207
Chishe Wang, Jiaxing Cheng, Shoubao Su, and Dongzhe Xu

Negative Generator Border for Effective Pattern Maintenance 217
Mengling Feng, Jinyan Li, Limsoon Wong, and Yap-Peng Tan

CommTracker: A Core-Based Algorithm of Tracking Community Evolution 229
Yi Wang, Bin Wu, and Xin Pei

Face Recognition Using Clustering Based Optimal Linear Discriminant Analysis 241
Wenxin Yang, Shuqin Rao, Jina Wang, Jian Yin, and Jian Chen

A Novel Immune Based Approach for Detection of Windows PE Virus 250
Yu Zhang, Tao Li, Jia Sun, and Renchao Qin

Using Genetic Algorithms for Parameter Optimization in Building Predictive Data Mining Models 260
Ashish Sureka and Kishore Varma Indukuri

Using Data Mining Methods to Predict Personally Identifiable Information in Emails 272
Liqiang Geng, Larry Korba, Xin Wang, Yunli Wang, Hongyu Liu, and Yonghua You

Iterative Reinforcement Cross-Domain Text Classification 282
Di Zhang, Gui-Rong Xue, and Yong Yu

Extracting Decision Rules from Sigmoid Kernel 294
Quanzhong Liu, Yang Zhang, and Zhengguo Hu

DMGrid: A Data Mining System Based on Grid Computing 305
Yi Wang, Liutong Xu, Guanhui Geng, Xiangang Zhao, and Nan Du

S-SimRank: Combining Content and Link Information to Cluster Papers Effectively and Efficiently 317
Yuanzhe Cai, Pei Li, Hongyan Liu, Jun He, and Xiaoyong Du

Open Domain Recommendation: Social Networks and Collaborative Filtering 330
Sarah K. Tyler and Yi Zhang

An Effective Approach for Identifying Evolving Three-Dimensional Structural Motifs in Protein Folding Data 342
Hui Yang and Lin Han

Texture Image Retrieval Based on Contourlet Transform and Active Perceptual Similarity Learning 355
Huaijing Qu, Yuhua Peng, Honglin Wan, and Min Han

A Temporal Dominant Relationship Analysis Method 367
Jing Yang, Yuanxi Wu, Cuiping Li, Hong Chen, and Bo Qu

Leakage-Aware Energy Efficient Scheduling for Fixed-Priority Tasks with Preemption Thresholds 379
XiaoChuan He and Yan Jia

Short Papers

Learning and Inferences of the Bayesian Network with Maximum Likelihood Parameters 391
JiaDong Zhang, Kun Yue, and WeiYi Liu

TARtool: A Temporal Dataset Generator for Market Basket Analysis 400
Asem Omari, Regina Langer, and Stefan Conrad

Dimensionality Reduction for Classification: Comparison of Techniques and Dimension Choice 411
Frank Plastria, Steven De Bruyne, and Emilio Carrizosa

Trajectories Mining for Traffic Condition Renewing 419
Danhuai Guo

Mining Bug Classifier and Debug Strategy Association Rules for Web-Based Applications 427
Lian Yu, Changzhu Kong, Lei Xu, Jingtao Zhao, and HuiHui Zhang

Test the Overall Significance of p-Values by Using Joint Tail Probability of Ordered p-Values as Test Statistic 435
Yongxiang Fang and Ernst Wit

Mining Interesting Infrequent and Frequent Itemsets Based on MLMS Model 444
Xiangjun Dong, Zhendong Niu, Donghua Zhu, Zhiyun Zheng, and Qiuting Jia

Text Learning and Hierarchical Feature Selection in Webpage Classification 452
Xiaogang Peng, Zhong Ming, and Haitao Wang

The RSO Algorithm for Reducing Number of Set Operations in Association Rule Mining 460
Muhammad Sarwar and Onaiza Maqbool

Predictive Performance of Clustered Feature-Weighting Case-Based Reasoning 469
Sung Ho Ha, Jong Sik Jin, and Jeong Won Yang

Selecting the Right Features for Bipartite-Based Text Clustering 477
Chao Qu, Yong Li, Jie Zhang, Tianming Hu, and Qian Chen

Image Emotional Classification Based on Color Semantic Description ... 485
Kaiping Wei, Bin He, Tao Zhang, and Wenya He

A Semi-supervised Clustering Algorithm Based on Must-Link Set 492
Haichao Huang, Yong Cheng, and Ruilian Zhao

T-rotation: Multiple Publications of Privacy Preserving Data Sequence 500
Youdong Tao, Yunhai Tong, Shaohua Tan, Shiwei Tang, and Dongqing Yang

The Integrated Methodology of KPCA and Wavelet Support Vector Machine for Predicting Financial Distress 508
Jian-guo Zhou, Tao Bai, and Ji-ming Tian

Outlier Detection Based on Voronoi Diagram 516
Jilin Qu

AWSum – Data Mining for Insight 524
Anthony Quinn, Andrew Stranieri, John Yearwood, and Gaudenz Hafen

Integrative Neural Network Approach for Protein Interaction Prediction from Heterogeneous Data 532
Xue-wen Chen, Mei Liu, and Yong Hu

Rules Extraction Based on Data Summarisation Approach Using DARA 540
Rayner Alfred

A Rough-Apriori Technique in Mining Linguistic Association Rules 548
Yun-Huoy Choo, Azuraliza Abu Bakar, and Abdul Razak Hamdan

Mining Causal Knowledge from Diagnostic Knowledge 556
Xiangdong An and Nick Cercone

Modified Particle Swarm Optimizer with Adaptive Dynamic Weights for Cancer Combinational Chemotherapy 563
Harish Chandra Soundararajan, Jagannathan Raman, and R. Muthucumaraswamy

MPSQAR: Mining Quantitative Association Rules Preserving Semantics 572
Chunqiu Zeng, Jie Zuo, Chuan Li, Kaikuo Xu, Shengqiao Ni, Liang Tang, Yue Zhang, and Shaojie Qiao

Using Support Vector Regression for Classification 581
Bo Huang, Zhihua Cai, Qiong Gu, and Changjun Chen

Dynamic Growing Self-organizing Neural Network for Clustering 589
Daxin Tian, Yueou Ren, and Qiuju Li

A Design of Reward Function Based on Knowledge in Multi-agent Learning 596
Bo Fan and Jiexin Pu

A Learning Method of Detecting Anomalous Pedestrian 604
Yue Liu, Jun Zhang, and Zhijing Liu

Moment+: Mining Closed Frequent Itemsets over Data Stream 612
Haifeng Li and Hong Chen

CDPM: Finding and Evaluating Community Structure in Social Networks 620
Li Wan, Jianxin Liao, and Xiaomin Zhu

Using Matrix Model to Find Association Rule Core for Diverse Compound Critiques 628
Li Yu

Link-Contexts for Ranking 636
Jessica Gronski

DC-Tree: An Algorithm for Skyline Query on Data Streams 644
Jing Yang, Bo Qu, Cui-Ping Li, and Hong Chen

Sequential Pattern Mining for Protein Function Prediction 652
Miao Wang, Xue-qun Shang, and Zhan-huai Li

Improving Web Search by Categorization, Clustering, and Personalization 659
Dengya Zhu and Heinz Dreher

JSNVA: A Java Straight-Line Drawing Framework for Network Visual Analysis 667
Qi Ye, Bin Wu, and Bai Wang

Recognition of Data Records in Semi-structured Web-Pages Using Ontology and χ^2 Statistical Distribution 675
Amin Keshavarzi, Amir Masoud Rahmani, Mehran Mohsenzadeh, and Reza Keshavarzi

Organizing Structured Deep Web by ClusteringQuery Interfaces Link Graph 683
Pengpeng Zhao, Li Huang, Wei Fang, and Zhiming Cui

CBP: A New Efficient Method for Mining Multilevel and Generalized Frequent Itemsets 691
Yu Xing Mao and Bai Le Shi

Supporting Customer Retention through Real-Time Monitoring of Individual Web Usage 699
Peter I. Hofgesang and Jan Peter Patist

A Comparative Study of Correlation Measurements for Searching Similar Tags 709
Kaikuo Xu, Yu Chen, Yexi Jiang, Rong Tang, Yintian Liu, and Jie Gong

Structure of Query Modification Process: Branchings 717
Nikolai Buzikashvili

Mining Top-n Local Outliers in Constrained Spatial Networks 725
Chongsheng Zhang, Zhongbo Wu, Bo Qu, and Hong Chen

Mining Concept-Drifting Data Streams with Multiple Semi-Random Decision Trees 733
Peipei Li, Xuegang Hu, and Xindong Wu

Automatic Web Tagging and Person Tagging Using Language Models 741
Qiaozhu Mei and Yi Zhang

Real-Time Person Tracking Based on Data Field 749
Shuliang Wang, Juebo Wu, Feng Cheng, and Hong Jin

Author Index 757