Lecture Notes in Artificial Intelligence 3584

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Springer Berlin

Berlin Heidelberg New York Hong Kong London Milan Paris Tokyo Xue Li Shuliang Wang Zhao Yang Dong (Eds.)

Advanced Data Mining and Applications

First International Conference, ADMA 2005 Wuhan, China, July 22-24, 2005 Proceedings



Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Xue Li University of Queensland School of Information Technology and Electrical Engineering Brisbane 4072, Queensland, Australia E-mail: xueli@itee.uq.edu.au

Shuliang Wang Wuhan University International School of Software Wuhan 430072, China E-mail: slwang2005@whu.edu.cn

Zhao Yang Dong University of Queensland School of Information Technology and Electrical Engineering St. Lucia, Queensland 4072, Australia E-mail: zdong@itee.uq.edu.au

Library of Congress Control Number: 2005929058

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

ISSN 0302-9743

ISBN-10 3-540-27894-X Springer Berlin Heidelberg New York ISBN-13 978-3-540-27894-8 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 springeronline.com

© Springer-Verlag Berlin Heidelberg 2005 Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India Printed on acid-free paper SPIN: 11527503 06/3142 5 4 3 2 1 0

Preface

With the ever-growing power to generate, transmit and collect 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 1st International Conference on Advanced Data Mining and Applications (ADMA 2005) 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 539 online submissions and 63 mailing submissions from 25 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 100 papers were accepted for the conference. Amongst them 25 papers were selected as regular papers and 75 papers were selected as short papers, yielding a combined acceptance rate of 17%.

The ADMA 2005 program highlights were four keynote speeches from outstanding researchers in advanced data mining and application areas: David Olson, Deyi Li, Chenqi Zhang, and Osmar Zaïane. The conference also invited researchers from two Australian universities to report on their latest research findings.

May 2005 Xue Li,
Shuliang Wang,
Zhaoyang Dong

Conference Committee

ADMA 2005 was organized by the International School of Software, Wuhan University, China and the School of Information Technology and Electrical Engineering, the University of Queensland, Australia; sponsored by the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China and the WISE (Web Information Systems Engineering, http://www.i-wise.org/) Society; and technically co-sponsored by the IEEE Queensland Section.

Organizing Committee

Conference Co-chairs

Honorary Chair Lotfi A. Zadeh (Berkeley University of

California, USA)

Jingnan Liu (Wuhan University, China) Deren Li (Wuhan University, China)

Xiaofang Zhou (University of Queensland,

Australia)

Program Co-chairs Xue Li (University of Queensland,

Australia)

Shuliang Wang (Secretary-General, Wuhan

University, China)

Zhaoyang Dong (University of Queensland,

Australia)

Organizing Chairs Yixin Zeng and Min Chen (Wuhan University,

China)

Publicity Chair Yanchun Zhang (Victoria University, Australia)

Local Advisory Committee

Liu Jingnan, Chen Shaofang
Li Wenxing, Huang Jin
Zhou Chuangbing

Wuhan University, China
Wuhan University, China
Wuhan University, China

Liu Haixin Technology Bureau of Hubei Province, China

Local Organizing Committee

Liu Fang, Yang Jing, Wi Zhou Xiaoming, Lin Bin, Zhu Guobing, Zheng Jing,

Sun Ling, Li Li

Wuhan University, China

Program Committee Members

Jesus Aguilar, Spain Viorel Ariton, Romania Michael Bain, Australia Jose Luis Balcazar, Spain Elena Baralis, Italy Petr Berka, Czech Republic Michael R. Berthold, Germany Fernando Berzal, Spain Fuling Bian, China Francesco Bonchi, Italy Jean-Francois Boulicaut, France Rui Camacho, Portugal Guoqing Chen, China Min Chen, China Krzysztof Cios, USA Bruno Cremilleux, France Luc Dehaspe, Belgium Kaichang Di, USA Floriana Esposito, Italy Marcus Gallagher, Australia Joao Gama, Portugal Dragan Gamberger, Croatia Jean-Gabriel Ganascia, France Junbin Gao, Australia Christophe Giraud-Carrier, USA Bart Goethals, Belgium Michael Frank Goodchild, USA Vladimir Gorodetsky, Russia Jiawei Han, USA Keqing He, China Yi Hong, Australia Andreas Hotho, Germany Zhanyi Hu, China Alípio Jorge, Portugal Mehmed Kantardzic, USA Eamonn Keogh, USA Adam Krzyzak, Canada Andrew Kusiak, USA Longin Jan Latecki, USA Andre Ponce Leao, Brazil Devi Li, China Qiaoyun Li, USA Qing Li, Hong Kong, China

Hui Lin, Hong Kong, China Xuemin Lin, Australia Wanquan Liu, Australia Yungang Liu, China Giuseppe Manco, Italy Rosa Meo, Italy Dunja Mladenic, Slovenia Iveta Mrazova, Czech Republic Olfa Nasraoui, USA Daniel Neagu, UK Claire Nedellec, France Mircea Neogita, New Zealand Arlindo Oliveira, Portugal David L. Olson, USA Yonghong Peng, UK Johann Petrak, Austria Pearl Pu, Switzerland Raghu Ramakrishnan, USA Jan Rauch, Czech Republic Zbigniew W. Ras, USA Cesar Rego, USA Christophe Rigotti, France Joseph Roure, Spain Juho Rousu, UK Celine Rouveirol, France Daniel Sanchez, Spain Yucel Saygin, Turkey Marc Sebban, France Giovanni Semeraro, Italy Seyed A. Shahrestani, Australia Wenzhong Shi, Hong Kong, China Andrzej Skowron, Poland Robert H. Sloan, USA Carlos Soares, Portugal Olga Stepankova, Czech Republic Ah-Hwee Tan, Singapore Kay Chen Tan, Singapore Kok Kiong Tan, Singapore Arthur Tay, Singapore Luis Torgo, Portugal Shusaku Tsumoto, Japan Brijesh Verma, Australia Ricardo Vilalta, USA

Paul Vitanyi, The Netherlands
Dianhui Wang, Australia
Ke Wang, Canada
Wei Wang, USA
Xinzhou Wang, China
Xizhao Wang, China
Marco Wiering, The Netherlands
Janet Wiles, Australia
Raymond Hau-San Wong, Hong Kong,
China
Dash Wu, Canada
Dongming Xu, Australia

Zijiang Yang, Canada
Jeffrey Xu Yu, Hong Kong, China
Philip S. Yu, USA
Osmar R. Zaïane, Canada
Gerson Zaverucha, Brazil
Sarah Zelikovitz, USA
Benjamin Zhan, USA
Shichao Zhang, Australia
Chenghu Zhou, China
Djamel A. Zighed, France
Blaz Zupan, Slovenia

External Reviewers

Mohsin Ali, Australia
Dingyi Chen, Australia
Xia Chen, China
Marian Craciun, Romania
Tomaz Curk, Slovenia
Yi Ding, Australia
Gongde Guo, UK
Zi Huang, Australia
Zheng Liu, Australia
Gregor Leban, Slovenia
Juggapong Natwichai, Australia

Daniel Neagu, UK Son Nghu, Australia Anisah Nizar, Australia Christoph Schmitz, Germany Dawei Song, Australia Gerd Stumme, Germany Xingzhi Sun, Australia Yidong Yuan, Australia Shuai Zhang, UK Junhua Zhao, Australia

Table of Contents

Keynote Papers

Decision Making with Uncertainty and Data Mining David L. Olson, Desheng Wu	1
Complex Networks and Networked Data Mining Deyi Li, Guisheng Chen, Baohua Cao	10
In-Depth Data Mining and Its Application in Stock Market Chengqi Zhang, Shichao Zhang	13
Relevance of Counting in Data Mining Tasks Osmar R. Zaïane	14
Invited Papers	
Term Graph Model for Text Classification Wei Wang, Diep Bich Do, Xuemin Lin	19
A Latent Usage Approach for Clustering Web Transaction and Building User Profile Yanchun Zhang, Guandong Xu, Xiaofang Zhou	31
Association Rules	
Mining Quantitative Association Rules on Overlapped Intervals Qiang Tong, Baoping Yan, Yuanchun Zhou	43
An Approach to Mining Local Causal Relationships from Databases Yang Bo He, Zhi Geng, Xun Liang	51
Mining Least Relational Patterns from Multi Relational Tables Siti Hairulnita Selamat, Mustafa Mat Deris, Rabiei Mamat, Zuriana Abu Bakar	59
Finding All Frequent Patterns Starting from the Closure Mohammad El-Hajj, Osmar R. Zaïane	67
Multiagent Association Rules Mining in Cooperative Learning Systems Reda Alhajj, Mehmet Kaya	75

VisAR: A New Technique for Visualizing Mined Association Rules Kesaraporn Techapichetvanich, Amitava Datta	88
An Efficient Algorithm for Mining Both Closed and Maximal Frequent Free Subtrees Using Canonical Forms Ping Guo, Yang Zhou, Jun Zhuang, Ting Chen, Yan-Rong Kang	96
Classification	
E-CIDIM: Ensemble of CIDIM Classifiers Gonzalo Ramos-Jiménez, José del Campo-Ávila, Rafael Morales-Bueno	108
Partially Supervised Classification – Based on Weighted Unlabeled Samples Support Vector Machine Zhigang Liu, Wenzhong Shi, Deren Li, Qianqing Qin	118
Mining Correlated Rules for Associative Classification Jian Chen, Jian Yin, Jin Huang	130
A Comprehensively Sized Decision Tree Generation Method for Interactive Data Mining of Very Large Databases Hyontai Sug	141
Using Latent Class Models for Neighbors Selection in Collaborative Filtering Xiaohua Sun, Fansheng Kong, Xiaobing Yang, Song Ye	149
A Polynomial Smooth Support Vector Machine for Classification YuBo Yuan, TingZhu Huang	157
Reducts in Incomplete Decision Tables Renpu Li, Dao Huang	165
Learning k-Nearest Neighbor Naive Bayes for Ranking Liangxiao Jiang, Harry Zhang, Jiang Su	175
One Dependence Augmented Naive Bayes Liangxiao Jiang, Harry Zhang, Zhihua Cai, Jiang Su	186
Clustering	
A Genetic k-Modes Algorithm for Clustering Categorical Data Guojun Gan, Zijiang Yang, Jianhong Wu	195

Table of Contents	XIII
A Fast Fuzzy Clustering Algorithm for Large-Scale Datasets Lukui Shi, Pilian He	203
Clustering with Noising Method Yongguo Liu, Yan Liu, Kefei Chen	209
Extracting the Representative Failure Executions via Clustering Analysis Based on Markov Profile Model Chengying Mao, Yansheng Lu	217
Improvement on the Approximation Bound for Fuzzy-Neural Networks Clustering Method with Gaussian Membership Function Weimin Ma, Guoqing Chen	225
Optimal Fuzzy Modeling Based on Minimum Cluster Volume Can Yang, Jun Meng	232
An Efficient Clustering Approach for Large Document Collections Bo Han, Lishan Kang, Huazhu Song	240
Clustering Categorical Data Using Coverage Density Hua Yan, Lei Zhang, Yi Zhang	248
Novel Algorithms	
A New Support Vector Machine for Data Mining Haoran Zhang, Xiaodong Wang, Changjiang Zhang, Xiuling Xu	256
The Infinite Polynomial Kernel for Support Vector Machine Degang Chen, Qiang He, Xizhao Wang	267
Routing Attribute Data Mining Based on Rough Set Theory Yanbing Liu, Hong Tang, Menghao Wang, Shixin Sun	276
A Novel Data Mining Method Based on Ant Colony Algorithm Weijin Jiang, Yusheng Xu, Yuhui Xu	284
Context-Sensitive Regression Analysis for Distributed Data Yan Xing, Michael G. Madden, Jim Duggan, Gerard J. Lyons	292
Customer Churn Prediction Using Improved One-Class Support Vector Machine Yu Zhao, Bing Li, Xiu Li, Wenhuang Liu, Shouju Ren	300

The Application of Adaptive Partitioned Random Search in Feature Selection Problem	
Xiaoyan Liu, Huaiqing Wang, Dongming Xu	307
Heuristic Scheduling of Concurrent Data Mining Queries Marek Wojciechowski, Maciej Zakrzewicz	315
Using Gap-Insensitive String Kernel to Detect Masquerading Chuanhuan Yin, Shengfeng Tian, Shaomin Mu	323
A New Method for Linear Ill-Posed Problems: Iteration Method by Rectifying Eigenvalue Yugang Tian, Peijun Shi, Xinzhou Wang, Kun Qin	331
Text Mining	
A Non-VSM kNN Algorithm for Text Classification Zhi-Hong Deng, Shi-Wei Tang	339
A Study on Text Clustering Algorithms Based on Frequent Term Sets Xiangwei Liu, Pilian He	347
An Improvement of Text Association Classification Using Rules Weights Xiao-Yun Chen, Yi Chen, Rong-Lu Li, Yun-Fa Hu	355
Word Segmentation and POS Tagging for Chinese Keyphrase Extraction Xiaochun Huang, Jian Chen, Puliu Yan, Xin Luo	364
Learning User Profiles from Text in e-Commerce Marco Degemmis, Pasquale Lops, Stefano Ferilli, Nicola Di Mauro, Teresa M.A. Basile, Giovanni Semeraro	370
Multimedia Mining	
Data Mining Based on Objects in Video Flow with Dynamic Background Cheng Zeng, Jiaheng Cao, Ying Fang, Pei Du	382
An Approach to Compressed Image Retrieval Based on JPEG2000 Framework	
Jianguo Tang, Wenyin Zhang, Chao Li	391
Target Segmentation and Feature Extraction for Undersea Image Based on Function Transformation	
Fuyuan Peng, Yan Tian, Xi Yu, Guohua Xu, Qian Xia	400

ART in Image Reconstruction with Narrow Fan-Beam Based on Data Minning	
Zhong Qu, Junhao Wen, Dan Yang, Ling Xu, Yu Wu	407
Digits Speech Recognition Based on Geometrical Learning Wenming Cao, Xiaoxia Pan, Shoujue Wang, Jing Hu	415
A Novel Information Hiding Technique for Remote Sensing Image Xianmin Wang, Zequn Guan, Chenhan Wu	423
Content-Based News Video Mining Junqing Yu, Yunfeng He, Shijun Li	431
Automatic Image Registration via Clustering and Convex Hull Vertices	
Matching Xiangyu Yu, Hong Sun	439
Fingerprint Image Segmentation Based on Gaussian-Hermite Moments Lin Wang, Hongmin Suo, Mo Dai	446
Sequential Data Mining and Time Series Mining	
HVSM: A New Sequential Pattern Mining Algorithm Using Bitmap Representation Shijie Song, Huaping Hu, Shiyao Jin	155
	455
HGA-COFFEE: Aligning Multiple Sequences by Hybrid Genetic Algorithm	
Li-fang Liu, Hong-wei Huo, Bao-shu Wang	464
Independent Component Analysis for Clustering Multivariate Time Series Data	
Edmond H.C. Wu, Philip L.H. Yu	474
Applying Fuzzy Neural Network to Intrusion Detection Based on Sequences of System Calls	
Guiling Zhang, Jizhou Sun	483
Web Mining	
Design and Implementation of Web Mining System Based on Multi-agent Wenbin Hu, Bo Meng	491
, <u> </u>	

A Novel Framework for Web Page Classification Using Two-Stage Neural Network Yunfeng Li, Yukun Cao, Qingsheng Zhu, Zhengyu Zhu	499
Fuzzy Evaluation of Hotel Websites Rob Law	507
Querying Web Images by Topic and Example Specification Methods Ching-Cheng Lee, Rashmi Prabhakara	515
The Research on Fuzzy Data Mining Applied on Browser Records Qingzhang Chen, Jianghong Han, Yungang Lai, Wenxiu He, Keji Mao	527
Discovering Conceptual Page Hierarchy of a Web Site from User Traversal History Xia Chen, Minqiang Li, Wei Zhao, Ding-Yi Chen	536
Biomedical Mining	
Bayesian Neural Networks for Prediction of Protein Secondary Structure Jianlin Shao, Dong Xu, Lanzhou Wang, Yifei Wang	544
PromPredictor: A Hybrid Machine Learning System for Recognition and Location of Transcription Start Sites in Human Genome Tao Li, Chuanbo Chen	552
Robust Ensemble Learning for Cancer Diagnosis Based on Microarray Data Classification Yonghong Peng	564
A Comprehensive Benchmark of the Artificial Immune Recognition System (AIRS) Lingjun Meng, Peter van der Putten, Haiyang Wang	575
An Analysis of Missing Data Treatment Methods and Their Application to Health Care Dataset Peng Liu, Elia El-Darzi, Lei Lei, Christos Vasilakis,	۳۵۹
Panagiotis Chountas, Wei Huang	583 591

Mining Interesting Association Rules in Medical Images Haiwei Pan, Jianzhong Li, Zhang Wei	598
Hybrid Feature Ranking for Proteins Classification Ricco Rakotomalala, Faouzi Mhamdi, Mourad Elloumi	610
Predicting Subcellular Localization of Proteins Using Support Vector Machine with N-Terminal Amino Composition Yan-fu Li, Juan Liu	618
Advanced Applications	
The Dynamic Character Curve Adjusting Model of Electric Load Based on Data Mining Theory Xiaoxing Zhang, Haijun Ren, Yuming Liu, Qiyun Cheng, Caixin Sun	626
Using Boosting Learning Method for Intrusion Detection Wu Yang, Xiao-Chun Yun, Yong-Tian Yang	634
RoleOf Relationship and Its Meta Model for Design Pattern Instantiation Chengwan He, Fei He, Keqing He, Jin Liu, Wenjie Tu	642
Automatic Inspection of Industrial Sheetmetal Parts with Single Non-metric CCD Camera Yongjun Zhang	654
An Advanced Implementation of a Distributed Control Scheme Based on LonWorks System over IP Networks Il-Joo Shim, Kyung-Bae Chang, Ki-Hyung Yu, Dong-Woo Cho, Kyoo-Dong Song, Gwi-Tae Park	662
Structural Damage Detection by Integrating Independent Component Analysis and Support Vector Machine Huazhu Song, Luo Zhong, Bo Han	670
An LZ78 Based String Kernel Ming Li, Ronan Sleep	678
Classifying Class and Finding Community in UML Metamodel Network Bin Liu, Deyi Li, Jin Liu, Fei He	690

An Adaptive Network Intrusion Detection Method Based on PCA and Support Vector Machines Xin Xu, Xuening Wang	696
Improved Grid Information Service Using the Idea of File-Parted Replication Jingwei Huang, Qingfeng Fan, Qiongli Wu, Yanxiang He	704
Dynamic Shape Modeling of Consumers' Daily Load Based on Data Mining Lianmei Zhang, Shihong Chen, Qiping Hu	712
A Study on the Mechanism of Virtual SAN-Based Spatial Data Storage with Double-Thoroughfare in Grid Jinsong Gao, Wen Zhang, Zequn Guan	720
A BP Neural Network Predictor Model for Desulfurizing Molten Iron Zhijun Rong, Binbin Dan, Jiangang Yi	728
A Flexible Report Architecture Based on Association Rules Mining Qiping Hu	736
Security and Privacy Issues	
Privacy Preserving Naive Bayes Classification Peng Zhang, Yunhai Tong, Shiwei Tang, Dongqing Yang	744
A Further Study on Inverse Frequent Set Mining Xia Chen, Maria Orlowska	753
Spatial Data Mining	
Mathematical Analysis of Classifying Convex Clusters Based on Support Functionals Xun Liang	761
Linear Belts Mining from Spatial Database with Mathematical Morphological Operators Min Wang, Jiancheng Luo, Chenghu Zhou	769
Spatial Information Multi-grid for Data Mining Zhenfeng Shao, Deren Li	

A Uniform Framework of 3D Spatial Data Model and Data Mining	
from the Model	
Peng-gen Cheng	785
Mining Standard Land Duice with Tangian Spline Dunction	
Mining Standard Land Price with Tension Spline Function	700
Hanning Yuan, Wenzhong Shi, Jiabing Sun	792
Streaming Data Mining	
Mining Recent Frequent Itemsets in Data Streams by Radioactively	
Attenuating Strategy	
Lifeng Jia, Zhe Wang, Chunguang Zhou, Xiujuan Xu	804
Lifting Jul, Zite Wang, Changaang Zitoa, Atajaan Aa	004
User Subjectivity in Change Modeling of Streaming Itemsets	
Vasudha Bhatnagar, Sarabjeet Kaur Kochhar	812
A Grid-Based Clustering Algorithm for High-Dimensional Data Streams	
Yansheng Lu, Yufen Sun, Guiping Xu, Gang Liu	824
Author Index	833

Table of Contents XIX