

Lecture Notes in Artificial Intelligence 3918

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Wee Keong Ng Masaru Kitsuregawa
Jianzhong Li Kuiyu Chang (Eds.)

Advances in Knowledge Discovery and Data Mining

10th Pacific-Asia Conference, PAKDD 2006
Singapore, April 9-12, 2006
Proceedings



Springer

Volume Editors

Wee Keong Ng

Nanyang Technological University, Centre for Advanced Information Systems
Nanyang Avenue, N4-B3C-14, 639798, Singapore
E-mail: awkng@ntu.edu.sg

Masaru Kitsuregawa

University of Tokyo, Institute of Industrial Science
4-6-1 Komaba, Meguro-Ku, Tokyo 153-8305, Japan
E-mail: kitsure@tkl.iis.u-tokyo.ac.jp

Jianzhong Li

Harbin Institute of Technology
Department of Computer Science and Engineering
Harbin, Heilongjiang, China
E-mail: lijzh@hit.edu.cn

Kuiyu Chang

Nanyang Technological University, School of Computer Engineering
Singapore 639798, Singapore
E-mail: kuiyu.chang@pmail.ntu.edu.sg

Library of Congress Control Number: 2006923003

CR Subject Classification (1998): I.2, H.2.8, H.3, H.5.1, G.3, J.1, K.4

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-540-33206-5 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-33206-0 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 2006
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11731139 06/3142 5 4 3 2 1 0

In Loving Memory of
Professor Hongjun Lu (1945 – 2005)

Preface

The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) is a leading international conference in the area of data mining and knowledge discovery. This year marks the tenth anniversary of the successful annual series of PAKDD conferences held in the Asia Pacific region. It was with pleasure that we hosted PAKDD 2006 in Singapore again, since the inaugural PAKDD conference was held in Singapore in 1997.

PAKDD 2006 continues its tradition of providing an international forum for researchers and industry practitioners to share their new ideas, original research results and practical development experiences from all aspects of KDD data mining, including data cleaning, data warehousing, data mining techniques, knowledge visualization, and data mining applications.

This year, we received 501 paper submissions from 38 countries and regions in Asia, Australasia, North America and Europe, of which we accepted 67 (13.4%) papers as regular papers and 33 (6.6%) papers as short papers. The distribution of the accepted papers was as follows: USA (17%), China (16%), Taiwan (10%), Australia (10%), Japan (7%), Korea (7%), Germany (6%), Canada (5%), Hong Kong (3%), Singapore (3%), New Zealand (3%), France (3%), UK (2%), and the rest from various countries in the Asia Pacific region.

The large number of papers was beyond our anticipation and we had to increase the Program Committee at the last minute in order to ensure that all papers went through a rigorous review process, without overloading the PC members. We are glad that most papers were reviewed by three PC members despite the tight schedule. We express herewith our deep appreciation to all PC members and the external reviewers for their arduous support in the review process.

PAKDD 2006 made several other progresses giving the conference series more visibility. For the first time, PAKDD workshops had formal proceedings published under Springer's Lecture Note series. The organizers of the four workshops, namely BioDM, KDLL, KDXD and WISI, put together very high-quality keynotes and workshop programs. We would like to express our gratitude to them for the tremendous efforts. PAKDD 2006 also introduced the best paper award in addition to the existing best student paper award(s). With the help of the Singapore Institute of Statistics (SIS) and the Pattern Recognition & Machine Intelligence Association (PREMIA) of Singapore, a data mining competition under the PAKDD flag was also organized for the first time. Last but not least, a one-day PAKDD School, similar to the one organized in PAKDD 2004, was held again this year.

PAKDD 2006 would not have been possible without the support of many people and organizations. We wish to thank the members of the Steering Committee for their invaluable suggestions and support throughout the organization process. We are grateful to the members of the Organizing Committee, who devoted much of their precious time to the conference arrangement. In the early stage of our conference preparation, we lost Hongjun Lu, who had helped us immensely in drafting our conference proposal. We have missed him dearly but would like to continue his inspiration to make PAKDD 2006 a success. We also deeply appreciate the generous financial support of Infocomm Development Authority of Singapore, the Lee

Foundation, the SPSS, the SAS Institute, the U.S. Air Force Office of Scientific Research, the Asian Office of Aerospace Research and Development, and the U.S. Army ITC-PAC Asian Research Office.

Last but not least, we want to thank all authors and all conference participants for their contribution and support. We hope all participants took this opportunity to share and exchange ideas with one another and enjoyed the conference.

April 2006

Masaru Kitsuregawa
Jianzhong Li
Ee-Peng Lim
Wee Keong Ng
Jaideep Srivastava

Organization

PAKDD 2006 Conference Committee

General Chairs

Ee-Peng Lim	Nanyang Technological University, Singapore
Hongjun Lu (Late)	HK University of Science and Technology, China
Jaideep Srivastava	University of Minnesota, USA

Program Chairs

Wee-Keong Ng	Nanyang Technological University, Singapore
Jiangzhong Li	Harbin Institute of Technology, China
Masaru Kitsuregawa	University of Tokyo, Japan

Workshop Chairs

Ah-Hwee Tan	Nanyang Technological University, Singapore
Huan Liu	Arizona State University, USA

Tutorial Chairs

Sourav Saha Bhowmick	Nanyang Technological University, Singapore
Osmar R. Zaiane	University of Alberta, Canada

Industrial Track Chair

Limsoon Wong	I2R, Singapore
--------------	----------------

PAKDD School Chair

Chew Lim Tan	National University of Singapore, Singapore
--------------	---

Publication Chair

Kuiyu Chang	Nanyang Technological University, Singapore
-------------	---

Panel Chairs

Wynne Hsu	National University of Singapore, Singapore
Bing Liu	University of Illinois at Chicago, USA

Local Arrangement Chairs

Bastion Arlene	Nanyang Technological University, Singapore
Vivekanand Gopalkrishnan	Nanyang Technological University, Singapore
Dion Hoe-Lian Goh	Nanyang Technological University, Singapore

Publicity and Sponsorship Chairs

Manoranjan Dash	Nanyang Technological University, Singapore
Jun Zhang	Nanyang Technological University, Singapore

PAKDD 2006 Steering Committee

Hiroshi Motoda (Chair)	Osaka University, Japan
David Cheung (Co-chair & Treasurer)	University of Hong Kong, China
Ho Tu Bao	Japan Advanced Institute of Science and Technology, Japan
Arbee L. P. Chen	National Chengchi University, Taiwan
Ming-Syan Chen	National Taiwan University, Taiwan
Jongwoo Jeon	Seoul National University, Korea
Masaru Kitsuregawa	Tokyo University, Japan
Rao Kotagiri	University of Melbourne, Australia
Huan Liu	Arizona State University, USA
Takao Terano	University of Tsukuba, Japan
Kyu-Young Whang	Korea Advanced Institute of Science and Technology, Korea
Graham Williams	ATO, Australia
Ning Zhong	Maebashi Institute of Technology, Japan
Chengqi Zhang	University of Technology Sydney, Australia

PAKDD 2006 Program Committee

Graham Williams	ATO, Australia
Warren Jin	Commonwealth Scientific and Industrial Research Organisation, Australia
Honghua Dai	Deakin University, Australia
Kok Leong Ong	Deakin University, Australia
David Taniar	Monash University, Australia
Vincent Lee	Monash University, Australia
Kai Ming Ting	Monash University, Australia
Richi Nayak	Queensland University of Technology, Australia
Vic Ciesielski	RMIT University, Australia
Vo Ngoc Anh	University of Melbourne, Australia
Rao Kotagiri	University of Melbourne, Australia
Achim Hoffmann	University of New South Wales, Australia
Xuemin Lin	University of New South Wales, Australia
Sanjay Chawla	University of Sydney, Australia
Douglas Newlands	University of Tasmania, Australia
Simeon J. Simoff	University of Technology, Sydney, Australia
Chengqi Zhang	University of Technology, Sydney, Australia
Doan B. Hoang	University of Technology, Sydney, Australia
Nicholas Cercone	Dalhousie University, Canada
Doina Precup	McGill University, Canada
Jian Pei	Simon Fraser University, Canada
Yiyu Yao	University of Regina, Canada
Zhihai Wang	Beijing Jiaotong University, China
Hai Zhuge	Chinese Academy of Sciences, China
Ada Waichee Fu	Chinese University of Hong Kong, China
Shuigeng Zhou	Fudan University, China
Aoying Zhou	Fudan University, China
Jiming Liu	Hong Kong Baptist University, China
Qiang Yang	Hong Kong University of Science and Technology, China

Zhi-Hua Zhou	Nanjing University, China
Xiaofeng Meng	Renmin University of China, China
Bo Zhang	Tsinghua University, China
David Cheung	University of Hong Kong, China
Joshua Z. Huang	University of Hong Kong, China
Djamel A. Zighed	University Lyon 2, France
Joel Quinqueton	University Montpellier, France
Thu Hoang	University Paris 5, France
Wai Lam	Chinese University of Hong Kong, Hong Kong, China
Wilfred Ng	University of Science and Technology, Hong Kong, China
Ajay B Pandey	Government of India, India
P. S. Sastry	Indian Institute of Science, Bangalore, India
Shyam Kumar Gupta	Indian Institute of Technology, Delhi, India
T. V. Prabhakar	Indian Institute of Technology, Kanpur, India
A. Balachandran	Persistent Systems, India
Aniruddha Pant	Persistent Systems, India
Dino Pedreschi	Università di Pisa, Italy
Tomoyuki Uchida	Hiroshima City University, Japan
Tetsuya Murai	Hokkaido University, Japan
Hiroki Arimura	Hokkaido University, Japan
Tetsuya Yoshida	Hokkaido University, Japan
Tu Bao Ho	JAIST, Japan
Van Nam Huynh	JAIST, Japan
Akira Shimazu	JAIST, Japan
Kenji Satou	JAIST, Japan
Takahira Yamaguchi	Keio University, Japan
Takashi Okada	Kwansei Gakuin University, Japan
Ning Zhong	Maebashi Institute of Technology, Japan
Hiroyuki Kawano	Nanzan University, Japan
Masashi Shimbo	Nara Institute of Science and Technology, Japan
Yuji Matsumoto	Nara Institute of Science and Technology, Japan
Seiji Yamada	National Institute of Informatics, Japan
Hiroshi Motoda	Osaka University, Japan
Shusaku Tsumoto	Shimane Medical University, Japan
Hiroshi Tsukimoto	Tokyo Denki University, Japan
Takao Terano	Tsukuba University, Japan
Takehisa Yairi	University of Tokyo, Japan
Yoon-Joon Lee	KAIST, Korea
Yang-Sae Moon	Kangwon National University, Korea
Sungzoon Cho	Seoul National University, Korea
Myung Won Kim	Soongsil University, Korea
Sang Ho Lee	Soongsil University, Korea
Myo Win Khin	University of Computer Studies, Myanmar
Myo-Myo Naing	University of Computer Studies, Myanmar
Patricia Riddle	University of Auckland, New Zealand
Eibe Frank	University of Waikato, New Zealand
Michael Mayo	University of Waikato, New Zealand
Szymon Jaroszewicz	Technical University of Szczecin, Poland
Andrzej Skowron	Warsaw University, Poland
Hung Son Nguyen	Warsaw University, Poland
Marzena Kryszkiewicz	Warsaw University of Technology, Poland
Ngoc Thanh Nguyen	Wroclaw University of Technology, Poland

Joao Gama	University of Porto, Portugal
Jinyan Li	Institute for Infocomm Research, Singapore
Lihui Chen	Nanyang Technological University, Singapore
Manoranjan Dash	Nanyang Technological University, Singapore
Siu Cheung Hui	Nanyang Technological University, Singapore
Daxin Jiang	Nanyang Technological University, Singapore
Daming Shi	Nanyang Technological University, Singapore
Aixin Sun	Nanyang Technological University, Singapore
Vivekanand Gopalkrishnan	Nanyang Technological University, Singapore
Sourav Bhownick	Nanyang Technological University, Singapore
Lipo Wang	Nanyang Technological University, Singapore
Wynne Hsu	National University of Singapore, Singapore
Dell Zhang	National University of Singapore, Singapore
Zehua Liu	Yokogawa Engineering Asia, Singapore
Ming-Syan Chen	National Taiwan University, Taiwan
Arbee L.P. Chen	National Chengchi University, Taiwan
San-Yih Hwang	National Sun Yat-Sen University, Taiwan
Chih-Jen Lin	National Taiwan University, Taiwan
Jirapun Daengdej	Assumption University, Thailand
Jonathan Lawry	University of Bristol, UK
Huan Liu	Arizona State University, USA
Minos Garofalakis	Intel Research Laboratories, USA
Tao Li	Florida International University, USA
Wenke Lee	Georgia Tech University, USA
Philip S. Yu	IBM T.J. Watson Research Center, USA
Se June Hong	IBM T.J. Watson Research Center, USA
Rong Jin	Michigan State University, USA
Pusheng Zhang	Microsoft Corporation, USA
Mohammed J. Zaki	Rensselaer Polytechnic Institute, USA
Hui Xiong	Rutgers University, USA
Tsau Young Lin	San Jose State University, USA
Aleksandar Lazarevic	United Technologies, USA
Jason T. L. Wang	New Jersey Institute of Technology, USA
Sam Y. Sung	South Texas University, USA
Roger Chiang	University of Cincinnati, USA
Bing Liu	University of Illinois at Chicago, USA
Vipin Kumar	University of Minnesota, USA
Xintao Wu	University of North Carolina at Charlotte, USA
Yan Huang	University of North Texas, USA
Xindong Wu	University of Vermont, USA
Guozhu Dong	Wright State University, USA
Thanh Thuy Nguyen	Hanoi University Technology, Vietnam
Ngoc Binh Nguyen	Hanoi University Technology, Vietnam
Tru Hoang Cao	Ho Chi Minh City University of Technology, Vietnam

PAKDD 2006 External Reviewers

Alexandre Termier	Lei Tang	Vineet Chaoji
Andre Carvalho	Li Peng	Weiqiang Kong
Atorn Nuniyagul	Lin Deng	Wenny Rahayu
Aysel Ozgur	Liqin Zhang	Wojciech Jaworski
Ben Mayer	Lizhuang Zhao	Xiangdong An
Benjarath Phoophakdee	Longbing Cao	Xiaobo Peng
Brian Harrington	Lu An	Xiaoming Wu
Cai Yunpeng	Magdiel Galan	Xingquan Zhu
Canh-Hao Nguyen	Marc Ma	Xiong Wang
Chengjun Liu	Masahiko Ito	Xuelong Li
Chiara Renso	Masayuki Okabe	Yan Zhao
Cho Siu-Yeung, David	Maurizio Atzori	Yang Song
Choi Koon Kau, Byron	Michail Vlachos	Yanchang Zhao
Christophe Rigotti	Minh Le Nguyen	Yaohua Chen
Daan He	Mirco Nanni	Yasufumi Takama
Dacheng Tao	Miriam Baglioni	Yi Ping Ke
Dang-Hung Tran	Mohammed Al Hasan	Ying Yang
Dexi Liu	Mugdha Khaladkar	Yong Ye
Dirk Arnold	Nitin Agarwal	Zhaochun Yu
Dong-Joo Park	Niyati Parikh	Zheng Zhao
Dongrong Wen	Nguyen Phu Chien	Zhenxing Qin
Dragoljub Pokrajac	Pedro Rodrigues	Zhiheng Huang
Duong Tuan Anh	Qiang Zhou	Zhihong Chong
Eric Eilertson	Qiankun Zhao	Zujun Shentu
Feng Chen	Qing Liu	
Feng Gao	Qinghua Zou	
Fosca Giannotti	Rohit Gupta	
Francesco Bonchi	Saeed Salem	
Franco Turini	Sai Moturu	
Gaurav Pandey	Salvatore Ruggieri	
Gour C. Karmakar	Salvo Rinzivillo	
Haoliang Jiang	Sangjun Lee	
Hiroshi Murata	Saori Kawasaki	
Ho Lam Lau	Sen Zhang	
Hongjian Fan	Shichao Zhang	
Hongxing He	Shyam Boriah	
Hui Xiong	Songtao Guo	
Hui Zhang	Spiros Papadimitriou	
James Cheng	Surendra Singhi	
Jaroslav Stepaniuk	Takashi Onoda	
Jianmin Li	Terry Griffin	
Jiaqi Wang	Thanh-Phuong Nguyen	
Jie Chen	Thai-Binh Nguyen	
Jing Tian	Thoai Nam	
Jiye Li	Tianming Hu	
Junilda Spirollari	Tony Abou-Assaleh	
Katherine G. Herbert	Tsuyoshi Murata	
Kozo Ohara	Tuan Trung Nguyen	
Lance Parson	Varun Chandola	

Sponsorship

We wish to thank the following organizations for their contributions to the success of this conference:



Air Force Office of Scientific Research,
Asian Office of Aerospace Research and Development



US Army ITC-PAC Asian Research Office



Infocomm Development Authority of Singapore



SPSS, Inc.



Embassy of the United States of America, Singapore

Table of Contents

Keynote Speech

Protection or Privacy? Data Mining and Personal Data <i>David J. Hand</i>	1
The Changing Face of Web Search <i>Prabhakar Raghavan</i>	11

Invited Speech

Data Mining for Surveillance Applications <i>Bhavani M. Thuraisingham</i>	12
--	----

Classification

A Multiclass Classification Method Based on Output Design <i>Qi Qiang, Qinming He</i>	15
Regularized Semi-supervised Classification on Manifold <i>Lianwei Zhao, Siwei Luo, Yanchang Zhao, Lingzhi Liao, Zhihai Wang</i>	20
Similarity-Based Sparse Feature Extraction Using Local Manifold Learning <i>Cheong Hee Park</i>	30
Generalized Conditional Entropy and a Metric Splitting Criterion for Decision Trees <i>Dan A. Simovici, Szymon Jaroszewicz</i>	35
RNBL-MN: A Recursive Naive Bayes Learner for Sequence Classification <i>Dae-Ki Kang, Adrian Silvescu, Vasant Honavar</i>	45
TRIPPER: Rule Learning Using Taxonomies <i>Flavian Vasile, Adrian Silvescu, Dae-Ki Kang, Vasant Honavar</i>	55
Using Weighted Nearest Neighbor to Benefit from Unlabeled Data <i>Kurt Driessens, Peter Reutemann, Bernhard Pfahringer, Claire Leschi</i>	60

Constructive Meta-level Feature Selection Method Based on Method Repositories <i>Hidenao Abe, Takahira Yamaguchi</i>	70
---	----

Ensemble Learning

Variable Randomness in Decision Tree Ensembles <i>Fei Tony Liu, Kai Ming Ting</i>	81
Further Improving Emerging Pattern Based Classifiers Via Bagging <i>Hongjian Fan, Ming Fan, Kotagiri Ramamohanarao, Mengxu Liu</i>	91
Improving on Bagging with Input Smearing <i>Eibe Frank, Bernhard Pfahringer</i>	97
Boosting Prediction Accuracy on Imbalanced Datasets with SVM Ensembles <i>Yang Liu, Aijun An, Xiangji Huang</i>	107

Clustering

DeLiClu: Boosting Robustness, Completeness, Usability, and Efficiency of Hierarchical Clustering by a Closest Pair Ranking <i>Elke Achtert, Christian Böhm, Peer Kröger</i>	119
Iterative Clustering Analysis for Grouping Missing Data in Gene Expression Profiles <i>Dae-Won Kim, Bo-Yeong Kang</i>	129
An EM-Approach for Clustering Multi-Instance Objects <i>Hans-Peter Kriegel, Alexey Pryakhin, Matthias Schubert</i>	139
Mining Maximal Correlated Member Clusters in High Dimensional Database <i>Lizheng Jiang, Dongqing Yang, Shiwei Tang, Xiuli Ma, Dehui Zhang</i>	149
Hierarchical Clustering Based on Mathematical Optimization <i>Le Hoai Minh, Le Thi Hoai An, Pham Dinh Tao</i>	160
Clustering Multi-represented Objects Using Combination Trees <i>Elke Achtert, Hans-Peter Kriegel, Alexey Pryakhin, Matthias Schubert</i>	174

Parallel Density-Based Clustering of Complex Objects <i>Stefan Brecheisen, Hans-Peter Kriegel, Martin Pfeifle</i>	179
Neighborhood Density Method for Selecting Initial Cluster Centers in K-Means Clustering <i>Yunming Ye, Joshua Zhexue Huang, Xiaojun Chen, Shuigeng Zhou, Graham Williams, Xiaofei Xu</i>	189
Uncertain Data Mining: An Example in Clustering Location Data <i>Michael Chau, Reynold Cheng, Ben Kao, Jackey Ng</i>	199

Support Vector Machines

Parallel Randomized Support Vector Machine <i>Yumao Lu, Vwani Roychowdhury</i>	205
ε -Tube Based Pattern Selection for Support Vector Machines <i>Dongil Kim, Sungzoon Cho</i>	215
Self-adaptive Two-Phase Support Vector Clustering for Multi-Relational Data Mining <i>Ping Ling, Yan Wang, Chun-Guang Zhou</i>	225
One-Class Support Vector Machines for Recommendation Tasks <i>Yasutoshi Yajima</i>	230

Text and Document Mining

Heterogeneous Information Integration in Hierarchical Text Classification <i>Huai-Yuan Yang, Tie-Yan Liu, Li Gao, Wei-Ying Ma</i>	240
FISA: Feature-Based Instance Selection for Imbalanced Text Classification <i>Aixin Sun, Ee-Peng Lim, Boualem Benatallah, Mahbub Hassan</i>	250
Dynamic Category Profiling for Text Filtering and Classification <i>Rey-Long Liu</i>	255
Detecting Citation Types Using Finite-State Machines <i>Minh-Hoang Le, Tu-Bao Ho, Yoshiteru Nakamori</i>	265

A Systematic Study of Parameter Correlations in Large Scale Duplicate Document Detection <i>Shaozhi Ye, Ji-Rong Wen, Wei-Ying Ma</i>	275
Comparison of Documents Classification Techniques to Classify Medical Reports <i>F.H. Saad, B. de la Iglesia, G.D. Bell</i>	285
XCLS: A Fast and Effective Clustering Algorithm for Heterogenous XML Documents <i>Richi Nayak, Sumei Xu</i>	292
Clustering Large Collection of Biomedical Literature Based on Ontology-Enriched Bipartite Graph Representation and Mutual Refinement Strategy <i>Illhoi Yoo, Xiaohua Hu</i>	303
Web Mining	
Level-Biased Statistics in the Hierarchical Structure of the Web <i>Guang Feng, Tie-Yan Liu, Xu-Dong Zhang, Wei-Ying Ma</i>	313
CLEOPATRA: Evolutionary Pattern-Based Clustering of Web Usage Data <i>Qiankun Zhao, Sourav S. Bhowmick, Le Gruenwald</i>	323
Extracting and Summarizing Hot Item Features Across Different Auction Web Sites <i>Tak-Lam Wong, Wai Lam, Shing-Kit Chan</i>	334
Clustering Web Sessions by Levels of Page Similarity <i>Caren Moraes Nichele, Karin Becker</i>	346
<i>i</i> WED: An Integrated Multigraph Cut-Based Approach for Detecting Events from a Website <i>Qiankun Zhao, Sourav S. Bhowmick, Aixin Sun</i>	351
Enhancing Duplicate Collection Detection Through Replica Boundary Discovery <i>Zhigang Zhang, Weijia Jia, Xiaoming Li</i>	361

Graph and Network Mining

Summarization and Visualization of Communication Patterns in a Large-Scale Social Network <i>Preetha Appan, Hari Sundaram, Belle Tseng</i>	371
---	-----

Patterns of Influence in a Recommendation Network <i>Jure Leskovec, Ajit Singh, Jon Kleinberg</i>	380
Constructing Decision Trees for Graph-Structured Data by Chunkingless Graph-Based Induction <i>Phu Chien Nguyen, Kouzou Ohara, Akira Mogi, Hiroshi Motoda, Takashi Washio</i>	390
Combining Smooth Graphs with Semi-supervised Classification <i>Xueyuan Zhou, Chunping Li</i>	400
Network Data Mining: Discovering Patterns of Interaction Between Attributes <i>John Galloway, Simeon J. Simoff</i>	410
 Association Rule Mining	
SGPM: Static Group Pattern Mining Using Apriori-Like Sliding Window <i>John Goh, David Taniar, Ee-Peng Lim</i>	415
Mining Temporal Indirect Associations <i>Ling Chen, Sourav S. Bhowmick, Jinyan Li</i>	425
Mining Top-K Frequent Closed Itemsets Is Not in APX <i>Chienwen Wu</i>	435
Quality-Aware Association Rule Mining <i>Laure Berti-Équille</i>	440
IMB3-Miner: Mining Induced/Embedded Subtrees by Constraining the Level of Embedding <i>Henry Tan, Tharam S. Dillon, Fedja Hadzic, Elizabeth Chang, Ling Feng</i>	450
Maintaining Frequent Itemsets over High-Speed Data Streams <i>James Cheng, Yiping Ke, Wilfred Ng</i>	462
Generalized Disjunction-Free Representation of Frequent Patterns with at Most k Negations <i>Marzena Kryszkiewicz</i>	468
Mining Interesting Imperfectly Sporadic Rules <i>Yun Sing Koh, Nathan Rountree, Richard O'Keefe</i>	473

Improved Negative-Border Online Mining Approaches <i>Ching-Yao Wang, Shian-Shyong Tseng, Tzung-Pei Hong</i>	483
Association-Based Dissimilarity Measures for Categorical Data: Limitation and Improvement <i>Si Quang Le, Tu Bao Ho, Le Sy Vinh</i>	493
Is Frequency Enough for Decision Makers to Make Decisions? <i>Shichao Zhang, Jeffrey Xu Yu, Jingli Lu, Chengqi Zhang</i>	499
<i>Ramp</i> : High Performance Frequent Itemset Mining with Efficient Bit-Vector Projection Technique <i>Shariq Bashir, Abdul Rauf Baig</i>	504
Evaluating a Rule Evaluation Support Method Based on Objective Rule Evaluation Indices <i>Hidenao Abe, Shusaku Tsumoto, Miho Ohsaki, Takahira Yamaguchi</i>	509

Bio-data Mining

Scoring Method for Tumor Prediction from Microarray Data Using an Evolutionary Fuzzy Classifier <i>Shinn-Ying Ho, Chih-Hung Hsieh, Kuan-Wei Chen, Hui-Ling Huang, Hung-Ming Chen, Shinn-Jang Ho</i>	520
Efficient Discovery of Structural Motifs from Protein Sequences with Combination of Flexible Intra- and Inter-block Gap Constraints <i>Chen-Ming Hsu, Chien-Yu Chen, Ching-Chi Hsu, Baw-Jhiune Liu</i>	530
Finding Consensus Patterns in Very Scarce Biosequence Samples from Their Minimal Multiple Generalizations <i>Yen Kaow Ng, Takeshi Shinohara</i>	540
Kernels on Lists and Sets over Relational Algebra: An Application to Classification of Protein Fingerprints <i>Adam Woźnica, Alexandros Kalousis, Melanie Hilario</i>	546
Mining Quantitative Maximal Hyperclique Patterns: A Summary of Results <i>Yaochun Huang, Hui Xiong, Weili Wu, Sam Y. Sung</i>	552

Outlier and Intrusion Detection

A Nonparametric Outlier Detection for Effectively Discovering Top-N Outliers from Engineering Data <i>Hongqin Fan, Osmar R. Zaïane, Andrew Foss, Junfeng Wu</i>	557
A Fast Greedy Algorithm for Outlier Mining <i>Zengyou He, Shengchun Deng, Xiaofei Xu, Joshua Zherue Huang</i>	567
Ranking Outliers Using Symmetric Neighborhood Relationship <i>Wen Jin, Anthony K.H. Tung, Jiawei Han, Wei Wang</i>	577
Construction of Finite Automata for Intrusion Detection from System Call Sequences by Genetic Algorithms <i>Kyubum Wee, Sinjae Kim</i>	594
An Adaptive Intrusion Detection Algorithm Based on Clustering and Kernel-Method <i>Hansung Lee, Yongwha Chung, Daihee Park</i>	603
Weighted Intra-transactional Rule Mining for Database Intrusion Detection <i>Abhinav Srivastava, Shamik Sural, A.K. Majumdar</i>	611

Privacy

On Robust and Effective K-Anonymity in Large Databases <i>Wen Jin, Rong Ge, Weining Qian</i>	621
Achieving Private Recommendations Using Randomized Response Techniques <i>Huseyin Polat, Wenliang Du</i>	637
Privacy-Preserving SVM Classification on Vertically Partitioned Data <i>Hwanjo Yu, Jaideep Vaidya, Xiaoqian Jiang</i>	647

Relational Database

Data Mining Using Relational Database Management Systems <i>Beibei Zou, Xuesong Ma, Bettina Kemme, Glen Newton, Doina Precup</i>	657
Bias-Free Hypothesis Evaluation in Multirelational Domains <i>Christine Körner, Stefan Wrobel</i>	668

Enhanced DB-Subdue: Supporting Subtle Aspects of Graph Mining
Using a Relational Approach
*Ramanathan Balachandran, Srihari Padmanabhan,
Sharma Chakravarthy* 673

Multimedia Mining

Multimedia Semantics Integration Using Linguistic Model
Bo Yang, Ali R. Hurson 679

A Novel Indexing Approach for Efficient and Fast Similarity Search of
Captured Motions
Chuanjun Li, B. Prabhakaran 689

Mining Frequent Spatial Patterns in Image Databases
Wei-Ta Chen, Yi-Ling Chen, Ming-Syan Chen 699

Image Classification Via LZ78 Based String Kernel: A Comparative
Study
Ming Li, Yanong Zhu 704

Stream Data Mining

Distributed Pattern Discovery in Multiple Streams
Jimeng Sun, Spiros Papadimitriou, Christos Faloutsos 713

COMET: Event-Driven Clustering over Multiple Evolving Streams
Mi-Yen Yeh, Bi-Ru Dai, Ming-Syan Chen 719

Variable Support Mining of Frequent Itemsets over Data Streams Using
Synopsis Vectors
Ming-Yen Lin, Sue-Chen Hsueh, Sheng-Kun Hwang 724

Hardware Enhanced Mining for Association Rules
Wei-Chuan Liu, Ken-Hao Liu, Ming-Syan Chen 729

A Single Index Approach for Time-Series Subsequence Matching That
Supports Moving Average Transform of Arbitrary Order
Yang-Sae Moon, Jinho Kim 739

Efficient Mining of Emerging Events in a Dynamic Spatiotemporal
Environment
Yu Meng, Margaret H. Dunham 750

Temporal Data Mining

A Multi-Hierarchical Representation for Similarity Measurement of Time Series <i>Xinqiang Zuo, Xiaoming Jin</i>	755
Multistep-Ahead Time Series Prediction <i>Haibin Cheng, Pang-Ning Tan, Jing Gao, Jerry Scripps</i>	765
Sequential Pattern Mining with Time Intervals <i>Yu Hirate, Hayato Yamana</i>	775
A Wavelet Analysis Based Data Processing for Time Series of Data Mining Predicting <i>Weimin Tong, Yijun Li, Qiang Ye</i>	780

Novel Algorithms

Intelligent Particle Swarm Optimization in Multi-objective Problems <i>Shinn-Jang Ho, Wen-Yuan Ku, Jun-Wun Jou, Ming-Hao Hung, Shinn-Ying Ho</i>	790
Hidden Space Principal Component Analysis <i>Weida Zhou, Li Zhang, Licheng Jiao</i>	801
Neighbor Line-Based Locally Linear Embedding <i>De-Chuan Zhan, Zhi-Hua Zhou</i>	806
Predicting Rare Extreme Values <i>Luis Torgo, Rita Ribeiro</i>	816
Domain-Driven Actionable Knowledge Discovery in the Real World <i>Longbing Cao, Chengqi Zhang</i>	821
Evaluation of Attribute-Aware Recommender System Algorithms on Data with Varying Characteristics <i>Karen H.L. Tso, Lars Schmidt-Thieme</i>	831

Innovative Applications

An Intelligent System Based on Kernel Methods for Crop Yield Prediction <i>A. Majid Awan, Mohd. Noor Md. Sap</i>	841
---	-----

A Machine Learning Application for Human Resource Data Mining Problem
Zhen Xu, Binheng Song 847

Towards Automated Design of Large-Scale Circuits by Combining Evolutionary Design with Data Mining
Shuguang Zhao, Mingying Zhao, Jun Zhao, Licheng Jiao..... 857

Mining Unexpected Associations for Signalling Potential Adverse Drug Reactions from Administrative Health Databases
Huidong Jin, Jie Chen, Chris Kelman, Hongxing He, Damien McAullay, Christine M. O’Keefe 867

Author Index 877