

Lecture Notes in Artificial Intelligence 4426

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Zhi-Hua Zhou Hang Li  
Qiang Yang (Eds.)

# Advances in Knowledge Discovery and Data Mining

11th Pacific-Asia Conference, PAKDD 2007  
Nanjing, China, May 22-25, 2007  
Proceedings

## Series Editors

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

## Volume Editors

Zhi-Hua Zhou  
Nanjing University  
National Lab for Novel Software Technology  
Hankou Road 22, Nanjing 210093, China  
E-mail: zhouzh@nju.edu.cn

Hang Li  
Microsoft Research Asia  
No. 49 Zhichun Road, Haidian District, Beijing, China 100080  
E-mail: hangli@microsoft.com

Qiang Yang  
Hong Kong University of Science and Technology  
Department of Computer Science and Engineering  
Clearwater Bay, Kowloon, Hong Kong, China  
E-mail: qyang@cs.ust.hk

Library of Congress Control Number: 2007923867

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 3-540-71700-5 Springer Berlin Heidelberg New York  
ISBN-13 978-3-540-71700-3 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](http://springer.com)

© Springer-Verlag Berlin Heidelberg 2007  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India  
Printed on acid-free paper SPIN: 12042982 06/3180 5 4 3 2 1 0

# Preface

The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) has been held every year since 1997. This year, the 11th in the series (PAKDD 2007), was held at Nanjing, China, May 22–25, 2007. PAKDD is a leading international conference in the area of data mining. It provides an international forum for researchers and industry practitioners to share their new ideas, original research results and practical development experiences from all KDD-related areas including data mining, machine learning, databases, statistics, data warehousing, data visualization, automatic scientific discovery, knowledge acquisition and knowledge-based systems.

This year we received a record number of submissions. We received 730 research papers from 29 countries and regions in Asia, Australia, North America, South America, Europe and Africa. The submitted papers went through a rigorous reviewing process. Every submission except very few was reviewed by three reviewers. Moreover, for the first time, PAKDD 2007 introduced a procedure of having an area chair supervise the review process of every submission. Thus, most submissions were reviewed by four experts. The Program Committee members were deeply involved in a highly engaging selection process with discussions among reviewers and area chairs. When necessary, additional expert reviews were sought. As a result, a highly selective few were chosen to be presented at the conference, including only 34 (4.66%) regular papers and 92 (12.6%) short papers in these proceedings.

The PAKDD 2007 program also included four workshops. They were a workshop on Data Mining for Biomedical Applications (BioDM 2007), a workshop on Data Mining for Business (DMBiz 2007), a workshop on High-Performance Data Mining and Applications (HPDMA 2007) and a workshop on Service, Security and Its Data Management Technologies in Ubi-Com (SSDU 2007). A data mining competition under the PAKDD flag was also organized for the second time after the first competition that was held in PAKDD 2006.

PAKDD 2007 would not have been successful 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 indebted to the area chairs, Program Committee members and external reviewers for their effort and engagement in providing a rich and rigorous scientific program for PAKDD 2007. We wish to express our gratitude to our General Workshop Chair Takashi Washio for selecting and coordinating the exciting workshops, to the Tutorial and PAKDD School Chair Graham Williams for coordinating the fruitful tutorials and school lecturers, to the Industrial Track Chair Joshua Z. Huang for handling industrial track papers, to the PAKDD Competition Chair Nathaniel Noriel for organizing the PAKDD Competition and to the distinguished keynote speakers and tutorial presenters for their

wonderful talks and lectures. We are also grateful to the Local Arrangement Chairs Yang Gao and Xianglin Fei as well as the Local Organizing Committee, whose great effort ensured the success of the conference.

We greatly appreciate the support from various institutions. The conference was organized by the LAMDA Group of Nanjing University, Nanjing, China, in cooperation with Nanjing University of Aeronautics and Astronautics, the Japanese Society for Artificial Intelligence, and the Singapore Institute of Statistics. It was sponsored by the National Natural Science Foundation of China (NSFC), Microsoft AdCenter Labs, NEC Labs China, Microsoft Research Asia (MSRA), Salford Systems and K.C. Wong Education Foundation.

We also 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 PAKDD 2007.

January 2007

Zhi-Hua Zhou  
Hang Li  
Qiang Yang

# Organization

## PAKDD 2007 Conference Committee

### Honorary Chairs

Hiroshi Motoda AFOSR/AOARD and Osaka University, Japan  
(Life-long member)  
Ruqian Lu Chinese Academy of Sciences, China

### General Chairs

David W.-L. Cheung University of Hong Kong, China  
Jian Lu Nanjing University, China

### Program Committee Chairs

Zhi-Hua Zhou Nanjing University, China  
Hang Li Microsoft Research Asia, China  
Qiang Yang HKUST, China

### Local Arrangements Chairs

Yang Gao Nanjing University, China  
Xianglin Fei Nanjing University, China

### Workshop Chair

Takashi Washio Osaka University, Japan

### Tutorial Chair and PAKDD School Chair

Graham Williams University of Canberra, Australia

### Industrial Chairs

Joshua Z. Huang University of Hong Kong, China  
Yunming Ye Harbin Institute of Technology, China

### PAKDD Competition Chair

Nathaniel Noriel Singapore Institute of Statistics, Singapore

### Publicity Chair

Kay Chen Tan National University of Singapore, Singapore

### Publication Chair

Yuan Jiang Nanjing University, China

**Web Chair**

Yang Yu Nanjing University, China

**Publication and Registration Secretary**

Xu-Ying Liu Nanjing University, China

**PAKDD 2007 Conference Steering Committee**

**Chairs**

David W.-L. Cheung University of Hong Kong, China  
Rao Kotagiri University of Melbourne, Australia

**Treasurer**

Graham Williams University of Canberra, Australia

**Members**

Arbee L. P. Chen National Chengchi University, Taiwan  
Ming-Syan Chen National Taiwan University, Taiwan  
Tu Bao Ho Japan Advanced Institute of Science and  
Technology, Japan  
Masaru Kitsuregawa Tokyo University, Japan  
Huan Liu Arizona State University, USA  
Ee-Peng Lim Nanyang Technological University, Singapore  
Hiroshi Motoda AFOSR/AOARD and Osaka University, Japan  
(Life-long member)  
Jaideep Srivastava University of Minnesota, USA  
Takao Terano Tokyo Institute of Technology, Japan  
Kyu-Young Whang Korea Advanced Institute of Science and  
Technology, Korea  
Chengqi Zhang University of Technology Sydney, Australia  
Ning Zhong Maebashi Institute of Technology, Japan

**PAKDD 2007 Program Committee**

**Chairs**

Zhi-Hua Zhou Nanjing University, China  
Hang Li Microsoft Research Asia, China  
Qiang Yang HKUST, China

**Area Chairs**

Naoki Abe IBM T.J. Watson Research Center, USA  
Phoebe Chen Deakin University, Australia

Zheng Chen	Microsoft Research Asia, China
Lee-Feng Chien	Academia Sinica, Taiwan
Eibe Frank	University of Waikato, New Zealand
João Gama	LIACC-University Porto, Portugal
Achim Hoffmann	The University of New South Wales, Australia
James Kwok	HKUST, China
Jinyan Li	Institute for Infocomm Research, Singapore
Charles X. Ling	University of Western Ontario, Canada
Huan Liu	Arizona State University, USA
Wee Keong Ng	Nanyang Technological University, Singapore
Jian Pei	Simon Fraser University, Canada
Fabio Roli	University of Cagliari, Italy
Takao Terano	Tokyo Institute of Technology, Japan
Kai Ming Ting	Monash University, Australia
Wei Wang	University of North Carolina at Chapel Hill, USA
Shichao Zhang	Guangxi Normal University, China
Zhongfei (Mark) Zhang	SUNY Binghamton, USA
Zijian Zheng	Microsoft, USA

## Members

Gagan Agrawal	Vic Ciesielski
David Albrecht	Diane Cook
Aijun An	Alfredo Cuzzocrea
Vo Ngoc Anh	Dao-Qing Dai
Chid Apte	Honghua Dai
Hiroki Arimura	Gautam Das
Michael W. Berry	Tamraparni Dasu
Michael Berthold	Ian Davidson
Steffen Bickel	Luc De Raedt
Hendrik Blockeel	Xiaoyong Du
Jean-Francois Boulicaut	Tina Eliassi-Rad
Ulf Brefeld	Tapio Elomaa
Rui Camacho	Andries Engelbrecht
Longbing Cao	Floriana Esposito
Tru Hoang Cao	Johannes Fürnkranz
Sanjay Chawla	Wei Fan
Arbee Chen	Ada Waichee Fu
Ming-Syan Chen	Dragan Gamberger
Shu-Ching Chen	Junbin Gao
Songcan Chen	Rayid Ghani
Yixin Chen	Fosca Giannotti
William K. Cheung	Aristides Gionis
Yiu-ming Cheung	Bart Goethals
Sungzoon Cho	Dimitrios Gunopulos



Shyam Kumar Gupta  
Jiawei Han  
Hermann Helbig  
Tu Bao Ho  
Thu Hoang  
Wynne Hsu  
Xiaohua Hu  
Jimmy Huang  
Jin Huang  
San-Yih Hwang  
Sanjay Jain  
Szymon Jaroszewicz  
Daxin Jiang  
Licheng Jiao  
Huidong Jin  
Rong Jin  
Ruoming Jin  
Alipio M. Jorge  
Hillol Kargupta  
George Karypis  
Hiroyuki Kawano  
Eamonn Keogh  
Boonserm Kijssirikul  
Myung Won Kim  
Masaru Kitsuregawa  
Rao Kotagiri  
Marzena Kryszkiewicz  
Ravi Kumar  
Vipin Kumar  
Wai Lam  
Nada Lavrac  
Jonathan Lawry  
Sang Ho Lee  
Vincent C S Lee  
Wee Sun Lee  
Yoon-Joon Lee  
Tze Yun Leong  
Chun-hung Li  
Gang Li  
Jianzhong Li  
Tao Li  
Xiao-Lin Li  
Xue Li  
Xuelong Li  
Andrew Lim

Em-Peng Lim  
Chih-Jen Lin  
Xuemin Lin  
Tie-Yan Liu  
Xiaohui Liu  
Woong-Kee Loh  
Chang-Tien Lu  
Jixin Ma  
Marco Maggini  
Yutaka Matsuo  
Michael Mayo  
Sameep Mehta  
Wagner Meira Jr.  
Xiaofeng Meng  
Rosa Meo  
Toshiro Minami  
Pabitra Mitra  
Yang-Sae Moon  
Shinichi Morishita  
Hiroshi Motoda  
S. Muthu Muthukrishnan  
Atsuyoshi Nakamura  
Richi Nayak  
Wilfred Ng  
Hung Son Nguyen  
Ngoc Thanh Nguyen  
Zaiqing Nie  
Kamal Nigam  
Tadashi Nomoto  
Zoran Obradovic  
Takashi Okada  
Salvatore Orlando  
Matthew Otey  
Satoshi Oyama  
Sankar K. Pal  
Yi Pan  
Dino Pedreschi  
Wen-Chih Peng  
Yonghong Peng  
Vincenzo Piuri  
Joel Quinqueton  
Naren Ramakrishnan  
Sanjay Ranka  
Patricia Riddle  
Asim Roy

P. S. Sastry  
Kenji Satou  
Joern Schneidewind  
Dou Shen  
Yi-Dong Shen  
Shengli Sheng  
Daming Shi  
Zhongzhi Shi  
Akira Shimazu  
Masashi Shimbo  
Arno Siebes  
Andrzej Skowron  
Myra Spiliopoulou  
Ashok N. Srivastava  
Jaideep Srivastava  
Aixin Sun  
Einoshin Suzuki  
Ah-Hwee Tan  
Chew Lim Tan  
Pang-Ning Tan  
Zhaohui Tang  
David Taniar  
Theeramunkong Thanaruk  
Hannu Toivonen  
Luis Torgo  
Ivor W. Tsang  
Ah Chung Tsoi  
Shusaku Tsumoto  
Tomoyuki Uchida  
Jeffrey D. Ullman  
Benjamin W. Wah  
Guoyin Wang  
Haixun Wang  
Hui Wang  
Jason T. L. Wang  
Lipo Wang  
Wenjia Wang  
Xufa Wang  
Zhihai Wang  
Graham Williams  
Limsoon Wong  
Rebecca Wright  
Xindong Wu

Xintao Wu  
Zhaohui Wu  
Hui Xiong  
Zhuoming Xu  
Takehisa Yairi  
Seiji Yamada  
Chunsheng Yang  
Hui Yang  
Min Yao  
Yiyu Yao  
Jieping Ye  
Wai Kiang Yeap  
Dit-Yan Yeung  
Jian Yin  
Tetsuya Yoshida  
Clement Yu  
Hwanjo Yu  
Jeffrey Xu Yu  
Jian Yu  
Philip S. Yu  
Bianca Zadrozny  
Mohammed Zaki  
Bo Zhang  
Changshui Zhang  
Chengqi Zhang  
Daoqiang Zhang  
Du Zhang  
Harry Zhang  
Junping Zhang  
Kang Zhang  
Mengjie Zhang  
Weixiong Zhang  
Xuegong Zhang  
Zili Zhang  
Ning Zhong  
Sheng Zhong  
Aoying Zhou  
Shuigeng Zhou  
Yan Zhou  
Xiaojin Zhu  
Xingquan Zhu  
Djamel Abdelakder Zighed

## PAKDD 2007 External Reviewers

Osman Abul  
Gulsah Altun  
Bill Andreopoulos  
Mafruz Zaman Ashrafi  
Keven Ates  
Francesca Barrientos  
Jeremy Besson  
Shyam Boriah  
Toon Calders  
Nicolas Cebbron  
Vineet Chaoji  
Bernard Chen  
Minmin Chen  
Yun Chi  
Massimo Coppola  
Dayong Deng  
Kevin Deronne  
Nguyen Luong Dong  
Jianfeng Du  
Nuno Escudeiro  
Mohamed Gaber  
Feng Gao  
Jing Gao  
Ling Guo  
Songtao Guo  
Zhen Guo  
Rohit Gupta  
Sam Han  
Jie Hao  
Mohammad Al Hasan  
Hongxing He  
Chi-Wei Hsu  
Haejin Hu  
Ruoyun Huang  
Tao Jiang  
Chris Kauffman  
Sung Jin Kim  
Pushpa Kumar  
Man Lan  
Sangjun Lee  
Sute Lei  
Guoliang Li  
Ming Li

Yong Li  
Yuanxiang Li  
Hong-Cheu Liu  
Jiang Liu  
Nianjun Liu  
Xu-Ying Liu  
Yang Liu  
Ying Liu  
Shijian Lu  
Claudio Lucchese  
Feifei Ma  
Yuval Marom  
Rodney Martin  
Alessio Micheli  
Ieva Mitasiunaite  
Mirco Nanni  
Minh Le Nguyen  
Ann Nicholson  
Zhengyu Niu  
Masayuki Okabe  
Takashi Onoda  
Matt Otey  
Nikunj Oza  
Gaurav Pandey  
Steve Pellicer  
Raffaele Perego  
Benjarath Phoophakdee  
Marcin Pluciński  
Huzefa Rangwala  
Pedro Rodrigues  
Saeed Salem  
Ron Van Schyndel  
Jouni Seppanen  
Claudio Silvestri  
Fabrizio Silvestri  
Gyorgy Simon  
Lay Ki Soon  
John Stutz  
Yasufumi Takama  
Peter Tischer  
Nikil Wale  
Qian Wan  
Raymond Wan

Dong Wang  
Richard Watson  
William Webber  
Jun Xu  
You Xu  
Ghim Eng Yap  
Hang Yu

Yang Yu  
Huaifeng Zhang  
Wei Zhang  
Chunying Zhao  
Min Zhao  
Yanchang Zhao

Organized by



Nanjing University



LAMDA Group

In cooperation with



Nanjing University of  
Aeronautics and  
Astronautics



The Japanese Society for  
Artificial Intelligence



Singapore Institute of  
Statistics

Sponsored by



National Natural Science  
Foundation of China



Microsoft adCenter Labs



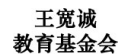
NEC Labs China



Microsoft Research Asia



Salford Systems



K.C. Wong Education  
Foundation

# Table of Contents

## Keynote Speeches

Research Frontiers in Advanced Data Mining Technologies and Applications . . . . .	1
<i>Jiawei Han</i>	
Finding the Real Patterns . . . . .	6
<i>Geoffrey Webb</i>	
Class Noise vs Attribute Noise: Their Impacts, Detection and Cleansing . . . . .	7
<i>Xindong Wu</i>	
Multi-modal and Multi-granular Learning . . . . .	9
<i>Bo Zhang and Ling Zhang</i>	

## Regular Papers

Hierarchical Density-Based Clustering of Categorical Data and a Simplification . . . . .	11
<i>Bill Andreopoulos, Aijun An, and Xiaogang Wang</i>	
Multi-represented Classification Based on Confidence Estimation . . . . .	23
<i>Johannes Aßfalg, Hans-Peter Kriegel, Alexey Pryakhin, and Matthias Schubert</i>	
Selecting a Reduced Set for Building Sparse Support Vector Regression in the Primal . . . . .	35
<i>Liefeng Bo, Ling Wang, and Licheng Jiao</i>	
Mining Frequent Itemsets from Uncertain Data . . . . .	47
<i>Chun-Kit Chui, Ben Kao, and Edward Hung</i>	
QC4 - A Clustering Evaluation Method . . . . .	59
<i>Daniel Crabtree, Peter Andrae, and Xiaoying Gao</i>	
Semantic Feature Selection for Object Discovery in High-Resolution Remote Sensing Imagery . . . . .	71
<i>Dihua Guo, Hui Xiong, Vijay Atluri, and Nabil Adam</i>	
Deriving Private Information from Arbitrarily Projected Data . . . . .	84
<i>Songtao Guo and Xintao Wu</i>	

Consistency Based Attribute Reduction . . . . .	96
<i>Qinghua Hu, Hui Zhao, Zongxia Xie, and Daren Yu</i>	
A Hybrid Command Sequence Model for Anomaly Detection . . . . .	108
<i>Zhou Jian, Haruhiko Shirai, Isamu Takahashi, Jousuke Kuroiwa, Tomohiro Odaka, and Hisakazu Ogura</i>	
$\sigma$ -Algorithm: Structured Workflow Process Mining Through Amalgamating Temporal Workcases . . . . .	119
<i>Kwanghoon Kim and Clarence A. Ellis</i>	
Multiscale BiLinear Recurrent Neural Network for Prediction of MPEG Video Traffic . . . . .	131
<i>Min-Woo Lee, Dong-Chul Park, and Yunsik Lee</i>	
An Effective Multi-level Algorithm Based on Ant Colony Optimization for Bisecting Graph . . . . .	138
<i>Ming Leng and Songnian Yu</i>	
A Unifying Method for Outlier and Change Detection from Data Streams Based on Local Polynomial Fitting . . . . .	150
<i>Zhi Li, Hong Ma, and Yongbing Mei</i>	
Simultaneous Tuning of Hyperparameter and Parameter for Support Vector Machines . . . . .	162
<i>Shizhong Liao and Lei Jia</i>	
Entropy Regularization, Automatic Model Selection, and Unsupervised Image Segmentation . . . . .	173
<i>Zhiwu Lu, Xiaoqing Lu, and Zhiyuan Ye</i>	
A Timing Analysis Model for Ontology Evolutions Based on Distributed Environments . . . . .	183
<i>Yinglong Ma, Beihong Jin, Yuancheng Li, and Kehe Wu</i>	
An Optimum Random Forest Model for Prediction of Genetic Susceptibility to Complex Diseases . . . . .	193
<i>Weidong Mao and Shannon Kelly</i>	
Feature Based Techniques for Auto-Detection of Novel Email Worms . . . . .	205
<i>Mohammad M. Masud, Latifur Khan, and Bhavani Thuraisingham</i>	
Multiresolution-Based BiLinear Recurrent Neural Network . . . . .	217
<i>Byung-Jae Min, Dong-Chul Park, and Hwan-Soo Choi</i>	
Query Expansion Using a Collection Dependent Probabilistic Latent Semantic Thesaurus . . . . .	224
<i>Laurence A.F. Park and Kotagiri Ramamohanarao</i>	

Scaling Up Semi-supervised Learning: An Efficient and Effective LLGC Variant .....	236
<i>Bernhard Pfahringer, Claire Leschi, and Peter Reutemann</i>	
A Machine Learning Approach to Detecting Instantaneous Cognitive States from fMRI Data .....	248
<i>Rafael Ramirez and Montserrat Puiggros</i>	
Discovering Correlated Items in Data Streams .....	260
<i>Xingzhi Sun, Ming Chang, Xue Li, and Maria E. Orlowska</i>	
Incremental Clustering in Geography and Optimization Spaces .....	272
<i>Chih-Hua Tai, Bi-Ru Dai, and Ming-Syan Chen</i>	
Estimation of Class Membership Probabilities in the Document Classification .....	284
<i>Kazuko Takahashi, Hiroya Takamura, and Manabu Okumura</i>	
A Hybrid Multi-group Privacy-Preserving Approach for Building Decision Trees .....	296
<i>Zhouxuan Teng and Wenliang Du</i>	
A Constrained Clustering Approach to Duplicate Detection Among Relational Data .....	308
<i>Chao Wang, Jie Lu, and Guangquan Zhang</i>	
Understanding Research Field Evolving and Trend with Dynamic Bayesian Networks .....	320
<i>Jinlong Wang, Congfu Xu, Gang Li, Zhenwen Dai, and Guojing Luo</i>	
Embedding New Data Points for Manifold Learning Via Coordinate Propagation .....	332
<i>Shiming Xiang, Feiping Nie, Yangqiu Song, Changshui Zhang, and Chunxia Zhang</i>	
Spectral Clustering Based Null Space Linear Discriminant Analysis (SNLDA) .....	344
<i>Wenxin Yang and Junping Zhang</i>	
On a New Class of Framelet Kernels for Support Vector Regression and Regularization Networks .....	355
<i>Wei-Feng Zhang, Dao-Qing Dai, and Hong Yan</i>	
A Clustering Algorithm Based on Mechanics .....	367
<i>Xianchao Zhang, He Jiang, Xinyue Liu, and Hong Yu</i>	
DLDA/QR: A Robust Direct LDA Algorithm for Face Recognition and Its Theoretical Foundation .....	379
<i>Yu-Jie Zheng, Zhi-Bo Guo, Jian Yang, Xiao-Jun Wu, and Jing-Yu Yang</i>	

gPrune: A Constraint Pushing Framework for Graph Pattern Mining . . . 388  
*Feida Zhu, Xifeng Yan, Jiawei Han, and Philip S. Yu*

**Short Papers**

Modeling Anticipatory Event Transitions . . . . . 401  
*Ridzwan Aminuddin, Ridzwan Suri, Kuiyu Chang, Zaki Zainudin, Qi He, and Ee-Peng Lim*

A Modified Relationship Based Clustering Framework for Density Based Clustering and Outlier Filtering on High Dimensional Datasets . . . . . 409  
*Turgay Tugay Bilgin and A. Yilmaz Camurcu*

A Region-Based Skin Color Detection Algorithm . . . . . 417  
*Faliang Chang, Zhiqiang Ma, and Wei Tian*

Supportive Utility of Irrelevant Features in Data Preprocessing . . . . . 425  
*Sam Chao, Yiping Li, and Mingchui Dong*

Incremental Mining of Sequential Patterns Using Prefix Tree . . . . . 433  
*Yue Chen, Jiankui Guo, Yaqin Wang, Yun Xiong, and Yangyong Zhu*

A Multiple Kernel Support Vector Machine Scheme for Simultaneous Feature Selection and Rule-Based Classification . . . . . 441  
*Zhenyu Chen and Jianping Li*

Combining Supervised and Semi-supervised Classifier for Personalized Spam Filtering . . . . . 449  
*Victor Cheng and Chun-hung Li*

Qualitative Simulation and Reasoning with Feature Reduction Based on Boundary Conditional Entropy of Knowledge . . . . . 457  
*Yusheng Cheng, Yousheng Zhang, Xuegang Hu, and Xiaoyao Jiang*

A Hybrid Incremental Clustering Method-Combining Support Vector Machine and Enhanced Clustering by Committee Clustering Algorithm . . . . . 465  
*Deng-Yiv Chiu and Kong-Ling Hsieh*

CCRM: An Effective Algorithm for Mining Commodity Information from Threaded Chinese Customer Reviews . . . . . 473  
*Huizhong Duan, Shenghua Bao, and Yong Yu*

A Rough Set Approach to Classifying Web Page Without Negative Examples . . . . . 481  
*Qiguo Duan, Duoqian Miao, and Kaimin Jin*



Evolution and Maintenance of Frequent Pattern Space When Transactions Are Removed . . . . .	489
<i>Mengling Feng, Guozhu Dong, Jinyan Li, Yap-Peng Tan, and Limsoon Wong</i>	
Establishing Semantic Relationship in Inter-query Learning for Content-Based Image Retrieval Systems . . . . .	498
<i>Chun Che Fung and Kien-Ping Chung</i>	
Density-Sensitive Evolutionary Clustering . . . . .	507
<i>Maoguo Gong, Licheng Jiao, Ling Wang, and Liefeng Bo</i>	
Reducing Overfitting in Predicting Intrinsically Unstructured Proteins . . . . .	515
<i>Pengfei Han, Xiuzhen Zhang, Raymond S. Norton, and Zhiping Feng</i>	
Temporal Relations Extraction in Mining Hepatitis Data . . . . .	523
<i>Tu Bao Ho, Canh Hao Nguyen, Saori Kawasaki, and Katsuhiko Takabayashi</i>	
Supervised Learning Approach to Optimize Ranking Function for Chinese FAQ-Finder . . . . .	531
<i>Guoping Hu, Dan Liu, Qingfeng Liu, and Ren-hua Wang</i>	
Combining Convolution Kernels Defined on Heterogeneous Sub-structures . . . . .	539
<i>Minlie Huang and Xiaoyan Zhu</i>	
Privacy-Preserving Sequential Pattern Release . . . . .	547
<i>Huidong Jin, Jie Chen, Hongxing He, and Christine M. O'Keefe</i>	
Mining Concept Associations for Knowledge Discovery Through Concept Chain Queries . . . . .	555
<i>Wei Jin, Rohini K. Srihari, and Xin Wu</i>	
Capability Enhancement of Probabilistic Neural Network for the Design of Breakwater Armor Blocks . . . . .	563
<i>Doo Kie Kim, Dong Hyawn Kim, Seong Kyu Chang, and Sang Kil Chang</i>	
Named Entity Recognition Using Acyclic Weighted Digraphs: A Semi-supervised Statistical Method . . . . .	571
<i>Kono Kim, Yephoon Yoon, Harksoo Kim, and Jungyun Seo</i>	
Contrast Set Mining Through Subgroup Discovery Applied to Brain Ischaemia Data . . . . .	579
<i>Petra Kralj, Nada Lavrač, Dragan Gamberger, and Antonija Krstačić</i>	

Intelligent Sequential Mining Via Alignment: Optimization Techniques for Very Large DB .....	587
<i>Hye-Chung Kum, Joong Hyuk Chang, and Wei Wang</i>	
A Hybrid Prediction Method Combining RBF Neural Network and FAR Model .....	598
<i>Yongle Lü and Rongling Lang</i>	
An Advanced Fuzzy C-Mean Algorithm for Regional Clustering of Interconnected Systems .....	606
<i>Sang-Hyuk Lee, Jin-Ho Kim, Se-Hwan Jang, Jong-Bae Park, Young-Hwan Jeon, and Sung-Yong Sohn</i>	
Centroid Neural Network with Bhattacharyya Kernel for GPDF Data Clustering .....	616
<i>Song-Jae Lee and Dong-Chul Park</i>	
Concept Interconnection Based on Many-Valued Context Analysis .....	623
<i>Yuxia Lei, Yan Wang, Baoxiang Cao, and Jiguo Yu</i>	
Text Classification for Thai Medicinal Web Pages .....	631
<i>Verayuth Lertnattee and Thanaruk Theeramunkong</i>	
A Fast Algorithm for Finding Correlation Clusters in Noise Data .....	639
<i>Jiuyong Li, Xiaodi Huang, Clinton Selke, and Jianming Yong</i>	
Application of Discrimination Degree for Attributes Reduction in Concept Lattice .....	648
<i>Ming Li and De-San Yang</i>	
A Language and a Visual Interface to Specify Complex Spatial Patterns .....	656
<i>Xiaohui Li and Yan Huang</i>	
Clustering Ensembles Based on Normalized Edges .....	664
<i>Yan Li, Jian Yu, Pengwei Hao, and Zhulin Li</i>	
Quantum-Inspired Immune Clonal Multiobjective Optimization Algorithm .....	672
<i>Yangyang Li and Licheng Jiao</i>	
Phase Space Reconstruction Based Classification of Power Disturbances Using Support Vector Machines .....	680
<i>Zhiyong Li and Weilin Wu</i>	
Mining the Impact Factors of Threads and Participators on Usenet Using Link Analysis .....	688
<i>Hongbo Liu, Jiaxin Wang, Yannan Zhao, and Zehong Yang</i>	

Weighted Rough Set Learning: Towards a Subjective Approach . . . . .	696
<i>Jinfu Liu, Qinghua Hu, and Daren Yu</i>	
Multiple Self-Splitting and Merging Competitive Learning Algorithm . . .	704
<i>Jun Liu and Kotagiri Ramamohanarao</i>	
A Novel Relative Space Based Gene Feature Extraction and Cancer Recognition . . . . .	712
<i>Xinguo Lu, Yaping Lin, Haijun Wang, Siwang Zhou, and Xiaolong Li</i>	
Experiments on Kernel Tree Support Vector Machines for Text Categorization . . . . .	720
<i>Ithipan Methasate and Thanaruk Theeramunkong</i>	
A New Approach for Similarity Queries of Biological Sequences in Databases . . . . .	728
<i>Hoong Kee Ng, Kang Ning, and Hon Wai Leong</i>	
Anomaly Intrusion Detection Based on Dynamic Cluster Updating . . . . .	737
<i>Sang-Hyun Oh and Won-Suk Lee</i>	
Efficiently Mining Closed Constrained Frequent Ordered Subtrees by Using Border Information . . . . .	745
<i>Tomonobu Ozaki and Takenao Ohkawa</i>	
Approximate Trace of Grid-Based Clusters over High Dimensional Data Streams . . . . .	753
<i>Nam Hun Park and Won Suk Lee</i>	
BRIM: An Efficient Boundary Points Detecting Algorithm . . . . .	761
<i>Bao-Zhi Qiu, Feng Yue, and Jun-Yi Shen</i>	
Syntactic Impact on Sentence Similarity Measure in Archive-Based QA System . . . . .	769
<i>Guang Qiu, Jiajun Bu, Chun Chen, Peng Huang, and Keke Cai</i>	
Semi-structure Mining Method for Text Mining with a Chunk-Based Dependency Structure . . . . .	777
<i>Issei Sato and Hiroshi Nakagawa</i>	
Principal Curves with Feature Continuity . . . . .	785
<i>Ming-ming Sun and Jing-yu Yang</i>	
Kernel-Based Linear Neighborhood Propagation for Semantic Video Annotation . . . . .	793
<i>Jinhui Tang, Xian-Sheng Hua, Yan Song, Guo-Jun Qi, and Xiuqing Wu</i>	
Learning Bayesian Networks with Combination of MRMR Criterion and EMI Method . . . . .	801
<i>Fengzhan Tian, Feng Liu, Zhihai Wang, and Jian Yu</i>	

A Cooperative Coevolution Algorithm of RBFNN for Classification . . . . .	809
<i>Jin Tian, Mingqiang Li, and Fuzan Chen</i>	
ANGEL: A New Effective and Efficient Hybrid Clustering Technique for Large Databases . . . . .	817
<i>Cheng-Fa Tsai and Chia-Chen Yen</i>	
Exploring Group Moving Pattern for an Energy-Constrained Object Tracking Sensor Network . . . . .	825
<i>Hsiao-Ping Tsai, De-Nian Yang, Wen-Chih Peng, and Ming-Syan Chen</i>	
ProMail: Using Progressive Email Social Network for Spam Detection . . . . .	833
<i>Chi-Yao Tseng, Jen-Wei Huang, and Ming-Syan Chen</i>	
Multidimensional Decision Support Indicator (mDSI) for Time Series Stock Trend Prediction . . . . .	841
<i>Kuralmani Vellaisamy and Jinyan Li</i>	
A Novel Support Vector Machine Ensemble Based on Subtractive Clustering Analysis . . . . .	849
<i>Cuiru Wang, Hejin Yuan, Jun Liu, Tao Zhou, and Huiling Lu</i>	
Keyword Extraction Based on PageRank . . . . .	857
<i>Jinghua Wang, Jianyi Liu, and Cong Wang</i>	
Finding the Optimal Feature Representations for Bayesian Network Learning . . . . .	865
<i>LiMin Wang, ChunHong Cao, XiongFei Li, and HaiJun Li</i>	
Feature Extraction and Classification of Tumor Based on Wavelet Package and Support Vector Machines . . . . .	871
<i>Shulin Wang, Ji Wang, Huowang Chen, and Shutao Li</i>	
Resource Allocation and Scheduling Problem Based on Genetic Algorithm and Ant Colony Optimization . . . . .	879
<i>Su Wang and Bo Meng</i>	
Image Classification and Segmentation for Densely Packed Aggregates . . . . .	887
<i>Weixing Wang</i>	
Mining Temporal Co-orientation Pattern from Spatio-temporal Databases . . . . .	895
<i>Ling-Yin Wei and Man-Kwan Shan</i>	
Incremental Learning of Support Vector Machines by Classifier Combining . . . . .	904
<i>Yi-Min Wen and Bao-Liang Lu</i>	

Clustering Zebrafish Genes Based on Frequent-Itemsets and Frequency Levels .....	912
<i>Daya C. Wimalasuriya, Sridhar Ramachandran, and Dejing Dou</i>	
A Practical Method for Approximate Subsequence Search in DNA Databases .....	921
<i>Jung-Im Won, Sang-Kyoon Hong, Jee-Hee Yoon, Sanghyun Park, and Sang-Wook Kim</i>	
An Information Retrieval Model Based on Semantics .....	932
<i>Chen Wu and Quan Zhang</i>	
AttributeNets: An Incremental Learning Method for Interpretable Classification .....	940
<i>Hu Wu, Yongji Wang, and Xiaoyong Huai</i>	
Mining Personalization Interest and Navigation Patterns on Portal .....	948
<i>Jing Wu, Pin Zhang, Zhang Xiong, and Hao Sheng</i>	
Cross-Lingual Document Clustering .....	956
<i>Ke Wu and Bao-Liang Lu</i>	
Grammar Guided Genetic Programming for Flexible Neural Trees Optimization .....	964
<i>Peng Wu and Yuehui Chen</i>	
A New Initialization Method for Clustering Categorical Data .....	972
<i>Shu Wu, Qingshan Jiang, and Joshua Zhexue Huang</i>	
L0-Constrained Regression for Data Mining .....	981
<i>Zhili Wu and Chun-hung Li</i>	
Application of Hybrid Pattern Recognition for Discriminating Paddy Seeds of Different Storage Periods Based on Vis/NIRS .....	989
<i>Li Xiaoli, Cao Fang, and He Yong</i>	
Density-Based Data Clustering Algorithms for Lower Dimensions Using Space-Filling Curves .....	997
<i>Bin Xu and Danny Z. Chen</i>	
Transformation-Based GMM with Improved Cluster Algorithm for Speaker Identification .....	1006
<i>Limin Xu, Zhenmin Tang, Keke He, and Bo Qian</i>	
Using Social Annotations to Smooth the Language Model for IR .....	1015
<i>Shengliang Xu, Shenghua Bao, Yong Yu, and Yunbo Cao</i>	
Affection Factor Optimization in Data Field Clustering .....	1022
<i>Hong Yang, Jianxin Liu, and Zhong Li</i>	

A New Algorithm for Minimum Attribute Reduction Based on Binary Particle Swarm Optimization with Vaccination . . . . .	1029
<i>Dongyi Ye, Zhaojiong Chen, and Jiankun Liao</i>	
Graph Nodes Clustering Based on the Commute-Time Kernel . . . . .	1037
<i>Luh Yen, Francois Fouss, Christine Decaestecker, Pascal Francq, and Marco Saerens</i>	
Identifying Synchronous and Asynchronous Co-regulations from Time Series Gene Expression Data . . . . .	1046
<i>Ying Yin, Yuhai Zhao, and Bin Zhang</i>	
A Parallel Algorithm for Learning Bayesian Networks . . . . .	1055
<i>Kui Yu, Hao Wang, and Xindong Wu</i>	
Incorporating Prior Domain Knowledge into a Kernel Based Feature Selection Algorithm . . . . .	1064
<i>Ting Yu, Simeon J. Simoff, and Donald Stokes</i>	
Geo-spatial Clustering with Non-spatial Attributes and Geographic Non-overlapping Constraint: A Penalized Spatial Distance Measure . . . . .	1072
<i>Bin Zhang, Wen Jun Yin, Ming Xie, and Jin Dong</i>	
GBKII: An Imputation Method for Missing Values . . . . .	1080
<i>Chengqi Zhang, Xiaofeng Zhu, Jilian Zhang, Yongsong Qin, and Shichao Zhang</i>	
An Effective Gene Selection Method Based on Relevance Analysis and Discernibility Matrix . . . . .	1088
<i>Li-Juan Zhang, Zhou-Jun Li, and Huo-Wang Chen</i>	
Towards Comprehensive Privacy Protection in Data Clustering . . . . .	1096
<i>Nan Zhang</i>	
A Novel Spatial Clustering with Obstacles Constraints Based on Particle Swarm Optimization and K-Medoids . . . . .	1105
<i>Xueping Zhang, Jiayao Wang, Mingguang Wu, and Yi Cheng</i>	
Online Rare Events Detection . . . . .	1114
<i>Jun Hua Zhao, Xue Li, and Zhao Yang Dong</i>	
Structural Learning About Independence Graphs from Multiple Databases	1122
<i>Qiang Zhao, Hua Chen, and Zhi Geng</i>	
An Effective Method For Calculating Natural Adjacency Relation in Spatial Database . . . . .	1131
<i>Renliang Zhao and Jiatian Li</i>	
K-Centers Algorithm for Clustering Mixed Type Data . . . . .	1140
<i>Wei-Dong Zhao, Wei-Hui Dai, and Chun-Bin Tang</i>	

Proposion and Analysis of a TCP Feature of P2P Traffic . . . . . 1148  
*Li-Juan Zhou, Zhi-Tang Li, and Hao Tu*

**Author Index** . . . . . 1157