

Lecture Notes in Artificial Intelligence 3518

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Tu Bao Ho David Cheung Huan Liu (Eds.)

Advances in Knowledge Discovery and Data Mining

9th Pacific-Asia Conference, PAKDD 2005
Hanoi, Vietnam, May 18-20, 2005
Proceedings

Series Editors

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

Volume Editors

Tu Bao Ho

Japan Advanced Institute of Science and Technology
1-1 Asahidai Tatsunokuchi, Ishikawa 923-1292, Japan
E-mail: bao@jaist.ac.jp

David Cheung

University of Hong Kong
Pokfulam Road, Hong Kong, China
E-mail: dcheung@csis.hku.hk

Huan Liu

Arizona State University
Tempe, AZ 85287-8809, USA
E-mail: hliu@asu.edu

Library of Congress Control Number: Applied for

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

ISSN 0302-9743

ISBN-10 3-540-26076-5 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-26076-9 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

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: 11430919 06/3142 5 4 3 2 1 0

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. 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, data warehousing, machine learning, databases, statistics, knowledge acquisition and automatic scientific discovery, data visualization, causality induction, and knowledge-based systems. This year's conference (PAKDD 2005) was the ninth of the PAKDD series, and carried the tradition in providing high-quality technical programs to facilitate research in knowledge discovery and data mining. It was held in Hanoi, Vietnam at the Melia Hotel, 18–20 May 2005.

We are pleased to provide some statistics about PAKDD 2005. This year we received 327 submissions (a 37% increase over PAKDD 2004), which is the highest number of submissions since the first PAKDD in 1997) from 28 countries/regions: Australia (33), Austria (1), Belgium (2), Canada (11), China (91), Switzerland (2), France (9), Finland (1), Germany (5), Hong Kong (11), Indonesia (1), India (2), Italy (2), Japan (21), Korea (51), Malaysia (1), Macau (1), New Zealand (3), Poland (4), Pakistan (1), Portugal (3), Singapore (12), Taiwan (19), Thailand (7), Tunisia (2), UK (5), USA (31), and Vietnam (9). The submitted papers went through a rigorous reviewing process. Each submission was reviewed by at least two reviewers, and most of them by three or four reviewers. The Program Committee members were deeply involved in a highly engaging selection process with discussions among reviewers, and, when necessary, additional expert reviews were sought. As a result, the PAKDD 2005 Program Committee accepted for publication and oral presentation 48 regular papers and 49 short papers, representing 14.6% and 14.9% acceptance rates, respectively. The PAKDD 2005 program also included two workshops (“Knowledge Discovery and Data Management in Biomedical Science” and “Rough Set Techniques in Knowledge Discovery”), and four tutorials (“Graph Mining Techniques and Their Applications,” “Rough Set Approach to KDD,” “Web Delta Mining: Opportunities and Solutions,” and “Advanced Techniques for Information and Image Classification for Knowledge Management and Decision Making”).

PAKDD 2005 would not have been possible without the help of many people and organizations. First and foremost, we would like to thank the members of the Steering Committee, the Program Committee and external reviewers for their invaluable contributions. We wish to express our gratitude to:

- Honorary conference chairs: Dang Vu Minh (President of the Vietnamese Academy of Science and Technology, Vietnam) and Hoang Van Phong (Minister of Science and Technology, Vietnam);
- Conference chairs: Hiroshi Motoda (Osaka University, Japan) and Phan Dinh Dieu (Vietnam National University, Hanoi, Vietnam);

- Keynote and invited speakers: Tom Mitchell (Carnegie Mellon University, USA), Nada Lavrac (J. Stefan Institute, Slovenia) and Unna Huh (Information and Communications University, Korea);
- Local organizing committee chairs: Luong Chi Mai (Institute of Information Technology, Hanoi, Vietnam) and Nguyen Ngoc Binh (Hanoi University of Technology, Vietnam);
- Workshop chair: Kyuseok Shim (National Korean University, Korea);
- Tutorial chair: Takashi Washio (Osaka University, Japan);
- Industrial chair: Wee Keong Ng (Nanyang Technological University, Singapore);
- Publicity chair: Tran Tuan Nam (Japan Advanced Institute of Science and Technology, Japan);
- Publication chair: Saori Kawasaki (Japan Advanced Institute of Science and Technology, Japan);
- Registration chair: Nguyen Trong Dung, Institute of Information Technology, Hanoi, Vietnam);
- Award selection committee: David Cheung (University of Hong Kong, China), Huan Liu (Arizona State University, USA) and Graham Williams (ATO, Australia);
- Chani Johnson for his tireless effort in supporting Microsoft's Conference Management Tool;
- Workshop organizers: Kenji Satou and Tu Bao Ho (Japan Advanced Institute of Science and Technology, Japan), Marcin S. Szczuka and Nguyen Hung Son (Warsaw University, Poland). Tutorialists: Sharma Chakravarthy (University of Texas at Arlington, USA), Sanjay Madria (University of Missouri-Rolla, USA), Nguyen Hung Son and Marcin S. Szczuka (Warsaw University, Poland) and Parag Kulkarni (Capsilon India, India).
- External reviewers.

We greatly appreciate the financial support from various sponsors: Japan Advanced Institute of Science and Technology (JAIST), Vietnamese Academy of Science and Technology (VAST), Ministry of Science and Technology of Vietnam (MoST), Hanoi University of Technology (HUT), AFOSR/AOARD, IBM and Oracle Vietnam.

Last but not least, we would like to thank all authors, and all conference attendees for their contribution and participation. Without them, we would not have had this conference. We hope all attendees took time to exchange ideas with each other and enjoyed PAKDD 2005.

May 2005

Tu Bao Ho, David Cheung, Huan Liu

PAKDD 2005 Steering Committee

Hiroshi Motoda (Chair)	Osaka University, Japan
David Cheung (Co-chair)	University of Hong Kong, China
Hongjun Lu (Treasurer)	Hong Kong University of Science & Technology, China
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
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 2005 Program Committee

Hiroki Arimura	Hokkaido University, Japan
Ho Tu Bao	Japan Advanced Institute of Science and Technology, Japan
Nguyen Ngoc Binh	Hanoi University Technology, Vietnam
Pavel Brazdil	University of Porto, Portugal
Tru Hoang Cao	Ho Chi Minh City University of Technology, Vietnam
Nicholas Cercone	Dalhousie University, Canada
Arbee L.P. Chen	National Chengchi University, Taiwan
Ming-Syan Chen	National Taiwan University, Taiwan
David Cheung	University of Hong Kong, China
Vic Ciesielski	RMIT University, Australia
Vincent Corruble	University of Paris 6, France
Jirapun Daengdej	Assumption University, Thailand
Honghua Dai	Deakin University, Australia
Manoranjan Dash	Nanyang Technological University, Singapore
AnHai Doan	University Illinois Urbana, USA
Guozhu Dong	Wright State University, USA
Nguyen Trong Dung	Institute of Information Technology, VAST, Vietnam
Peter A. Flach	University of Bristol, UK
Eibe Frank	University of Waikato, New Zealand
Joao Gama	University of Porto, Portugal
Minos Garofalakis	Bell Laboratories, USA
Sudipto Guha	University of Pennsylvania, USA
Dimitrios Gunopulos	University of California, Riverside, USA
Shyam Kumar Gupta	Indian Institute of Technology, Delhi, India
Peter Haddawy	Asian Institute of Technology, Thailand

Jiawei Han	University of Illinois, Urbana-Champaign, USA
Doan B. Hoang	University of Technology, Sydney, Australia
Thu Hoang	University of Paris 5, France
Achim Hoffmann	University of New South Wales, Australia
Se June Hong	IBM T.J. Watson Research Center, USA
Wynne Hsu	National University of Singapore, Singapore
Joshua Z. Huang	University of Hong Kong, China
Siu Cheung Hui	Nanyang Technological University, Singapore
San-Yih Hwang	National Sun Yat-Sen University, Taiwan
Jongwoo Jeon	Seoul National University, Korea
Rong Jin	Michigan State University, USA
Hiroyuki Kawano	Nanzan University, Japan
Gabriele Kern-Isberner	University of Dortmund, Germany
Hoang Kiem	Vietnam National University HCM, Vietnam
Boonserm Kijssirikul	Chulalongkorn University, Thailand
Myoung Ho Kim	Korea Advanced Institute of Science and Technology, Korea
Yasuhiko Kitamura	Kwansei Gakuin University, Japan
Masaru Kitsuregawa	University of Tokyo, Japan
Rao Kotagiri	University of Melbourne, Australia
Marzena Kryszkiewicz	Warsaw University of Technology, Poland
Vipin Kumar	University of Minnesota, USA
Jonathan Lawry	University of Bristol, UK
Aleksandar Lazarevic	University of Minnesota, USA
Doheon Lee	Korea Advanced Institute of Science and Technology, Korea
Geuk Lee	Hannam University, Korea
Kwang Hyung Lee	Korea Advanced Institute of Science and Technology, Korea
Sang Ho Lee	Soongsil University, Korea
Yoon-Joon Lee	Korea Advanced Institute of Science and Technology, Korea
Jinyan Li	Institute for Infocomm Research, Singapore
Tsau Young Lin	San Jose State University, USA
Bing Liu	University of Illinois at Chicago, USA
Huan Liu	Arizona State University, USA
Hongjun Lu	Hong Kong University of Science and Technology, China
Luong Chi Mai	Vietnamese Academy of Science and Technology, Vietnam
Yuji Matsumoto	Nara Institute of Science and Technology, Japan
Hiroshi Motoda	Osaka University, Japan
Tetsuya Murai	Hokkaido University, Japan
Yoshiteru Nakamori	Japan Advanced Institute of Science and Technology, Japan
Huynh Van Nam	Japan Advanced Institute of Science and Technology, Japan

Douglas Newlands	Deakin University, Australia
Wee Keong Ng	Nanyang Technological University, Singapore
Zaiqing Nie	Microsoft Research Asia, China
Monique Noirhomme	University of Notre Dame de la Paix, Belgium
Masayuki Numao	Osaka University, Japan
Takashi Okada	Kwansei Gakuin University, Japan
Dino Pedreschi	Università di Pisa, Italy
T.V. Prabhakar	Indian Institute of Technology Kanpur, India
Joel Quinqueton	University of Montpellier 2, France
Rajeev Rastogi	Bell Laboratories, USA
Kenji Satou	Japan Advanced Institute of Science and Technology, Japan
Michele Sebag	University of Paris, Orsay, France
Rudy Setiono	National University of Singapore, Singapore
Kyuseok Shim	Seoul National University, Korea
Akira Shimazu	Japan Advanced Institute of Science and Technology, Japan
Masashi Shimbo	Nara Institute of Science and Technology, Japan
Simeon J. Simoff	University of Technology, Sydney, Australia
Andrzej Skowron	Warsaw University, Poland
Nguyen Hung Son	Warsaw University, Poland
Takao Terano	Tsukuba University, Japan
Nguyen Thanh Thuy	Hanoi University Technology, Vietnam
Hiroshi Tsukimoto	Tokyo Denki University, Japan
Shusaku Tsumoto	Shimane Medical University, Japan
Anh Vo	University of Melbourne, Australia
Zhi-Hai Wang	Beijing Jiaotong University, China
Takashi Washio	Osaka University, Japan
Kyu-Young Whang	Korea Advanced Institute of Science and Technology, Korea
Graham Williams	ATO, Australia
Xindong Wu	University of Vermont, USA
Takehisa Yairi	University of Tokyo, Japan
Seiji Yamada	National Institute of Informatics, Japan
Takahira Yamaguchi	Keio University, Japan
Yiyu Yao	University of Regina, Canada
Tetsuya Yoshida	Hokkaido University, Japan
Philip S. Yu	IBM T.J. Watson Research Center, USA
Mohammed J. Zaki	Rensselaer Polytechnic Institute, USA
Chengqi Zhang	University of Technology, Sydney, Australia
Bo Zhang	Tsinghua University, China
Ning Zhong	Maebashi Institute of Technology, Japan
Zhi-Hua Zhou	Nanjing University, China
Djamel A. Zighed	University of Lyon 2, France

PAKDD 2005 External Reviewers

Alexandre Termier	Karlton Sequeira	Noboru Nakajima
Arkadiusz Wojna	Kouzou Ohara	Noriko Imafuji
Asad Satti	Kun-Ta Chuang	P. Rodrigues
Aysel Ozgur	Kwoh Chee Keong	Pabitra Mitra
Benjarath Bphhpakdee	Lance Parsons	Phu Chien Nguyen
Bi-Ru Dai	Le Anh Cuong	Pusheng Zhang
Carlos Pinto	Le Minh Hoang	Remco Bouckaert
Chen Chen	Le Si Quang	Rohit Gupta
Chiara Renso	Lei Tang	Ronaldo Prati
Chung-Wen Cho	Lei Yu	Salvatore Ruggieri
Daan He	Levent Ertoz	Sangkyum Kim
Dae-Won Kim	Li Lin	Shichao Zhang
Dayang Iskandar	Ling Zhuang	Shuigeng Zhou
Ding-Ying Chiu	Lizhuang Zhang	Shyam Boriah
Eamonn Keogh	Lizhuang Zhao	Steffi Soo
Ellery Chen	Mark Hsieh	Surendra Singhi
Feng Gao	Masahi Toyoda	Tadashi Ohmori
Francesco Bonchi	Masaki Kuremtasu	Takeshi Sagara
Gang Li	Maurizio Atzori	Varun Chandola
Gaurav Pandey	Michael Steinbach	Vic Ciesielski
Georges Koepfler	Miho Ohsaki	Vincent S.M. Tseng
Graham Cormode	Mikihiko Mori	Vineet Chaoji
		Vivekanand
Gulisong Nansierding	Milton Silva	Gopalkrishnan
Gyorgy Simon	Min-Ling Zhang	Vlado Keselj
Han Liu	Mintz Hsieh	Vu Van Thinh
Hidenao Abe	Mirco Nanni	Xiaolei Li
Ho Wai Shing	Miriam Baglioni	Xifeng Yan
Hui Xiong	Miyuki Nakano	Xingquan Zhu
Hui Zhang	Mohammed Zaki	Yasufumi Takama
Hung-Chen Chen	Muneaki Ohshima	Yen-Kuang Lu
I-Jen Chiang	Nagender Parimi	Yi Liu
Iko Pramudiono	Naoki Fukuta	Yi Xia
J. Gama	Narendra S. Chaudhari	Yifan Li
Jean-Daniel Zucker	Nenad Stankovic	Ying Lu
Jeng-Kuen Chiu	Ng Wil Lie	Yitong Wang
Jia Hu	Nguyen Canh Hao	Yongdai Kim
Jiri Navratil	Nguyen Duc Dung	Yun Chi
Jiye Li	Nguyen Le Minh	Yunxiao Ma
Juliana Hsieh	Nguyen Minh	Zheng Shao
Justin Zobel	Nguyen Thi Minh Hai	Zheng Zhao

In Loving Memory of Professor Hongjun Lu 1945–2005

Professor Lu was one of the founders of the PAKDD conference series. He played a key leadership role in nurturing and establishing the PAKDD conferences to become a world-recognized forum. He served as the Steering Committee Co-chair (1998–2001), and as Chair (2001–2003) of PAKDD. He was the Program Co-chair of the inaugural PAKDD (1997). He was honored with the inaugural PAKDD Distinguished Contribution Award (2005) for his significant and ongoing contributions in research and services to the advancement of the PAKDD community and series of conferences.

Professor Lu also served in many important and influential positions in the research community. He was elected as a Trustee of the VLDB Endowment in 2000. He was a member of the Advisory Board of ACM SIGMOD (1998–2002). He was an Editor for IEEE Transactions on Knowledge and Data Engineering (TKDE) (1996–2000) and for Knowledge and Information Systems: An International Journal (1998–2001). He has served on the program committees of numerous international conferences in databases.

Professor Lu passed away on March 3 from complications arising from his treatment for cancer. His research has made an impact in many areas, especially in the many important issues related to query processing and optimization, data warehousing and data mining. His long-term contributions through over 200 research publications in scientific journals, conferences and workshop proceedings have provided the foundations for many other researchers, and will be an ongoing contribution to our scientific endeavors for many years to come.

He will always be remembered as a great scholar, researcher, teacher and leader, and as a caring, considerate and compassionate friend to very many.

Table of Contents

Keynote Speech and Invited Talks

Machine Learning for Analyzing Human Brain Function <i>Tom Mitchell</i>	1
Subgroup Discovery Techniques and Applications <i>Nada Lavrač</i>	2
IT Development in the 21 st Century and Its Implications <i>Unna Huh</i>	15

Theoretic Foundations

Data Mining of Gene Expression Microarray via Weighted Prefix Trees <i>Tran Trang, Nguyen Cam Chi, Hoang Ngoc Minh</i>	21
Automatic Extraction of Low Frequency Bilingual Word Pairs from Parallel Corpora with Various Languages <i>Hiroshi Echizen-ya, Kenji Araki, Yoshio Mornouchi</i>	32
A Kernel Function Method in Clustering <i>Ling Zhang, Tao Wu, Yanping Zhang</i>	38
Performance Measurements for Privacy Preserving Data Mining <i>Nan Zhang, Wei Zhao, Jianer Chen</i>	43
Extraction of Frequent Few-Overlapped Monotone DNF Formulas with Depth-First Pruning <i>Yoshikazu Shima, Kouichi Hirata, Masateru Harao</i>	50

Association Rules

Rule Extraction from Trained Support Vector Machines <i>Ying Zhang, HongYe Su, Tao Jia, Jian Chu</i>	61
Pruning Derivative Partial Rules During Impact Rule Discovery <i>Shiyong Huang, Geoffrey I. Webb</i>	71
IGB: A New Informative Generic Base of Association Rules <i>Gh. Gasmı, S. Ben Yahia, E. Mephu Nguifo, Y. Slimani</i>	81

A Divide and Conquer Approach for Deriving Partially Ordered Sub-structures <i>Sadok Ben Yahia, Yahya Slimani, Jihen Rezgui</i>	91
Finding Sporadic Rules Using Apriori-Inverse <i>Yun Sing Koh, Nathan Rountree</i>	97
Automatic View Selection: An Application to Image Mining <i>Manoranjan Dash, Deepak Kolippakkam</i>	107
Pushing Tougher Constraints in Frequent Pattern Mining <i>Francesco Bonchi, Claudio Lucchese</i>	114
An Efficient Compression Technique for Frequent Itemset Generation in Association Rule Mining <i>Mafruz Zaman Ashrafi, David Taniar, Kate Smith</i>	125
Mining Time-Profiled Associations: An Extended Abstract <i>Jin Soung Yoo, Pusheng Zhang, Shashi Shekhar</i>	136
Online Algorithms for Mining Inter-stream Associations from Large Sensor Networks <i>K. K. Loo, Ivy Tong, Ben Kao</i>	143
Mining Frequent Ordered Patterns <i>Zhi-Hong Deng, Cong-Rui Ji, Ming Zhang, and Shi-Wei Tang</i>	150
 Biomedical Domains	
Conditional Random Fields for Transmembrane Helix Prediction <i>Lior Lukov, Sanjay Chawla, W. Bret Church</i>	155
A DNA Index Structure Using Frequency and Position Information of Genetic Alphabet <i>Woo-Cheol Kim, Sanghyun Park, Jung-Im Won, Sang-Wook Kim, Jee-Hee Yoon</i>	162
An Automatic Unsupervised Querying Algorithm for Efficient Information Extraction in Biomedical Domain <i>Min Song, Il-Yeol Song, Xiaohua Hu, Robert B. Allen</i>	173
Voting Fuzzy k-NN to Predict Protein Subcellular Localization from Normalized Amino Acid Pair Compositions <i>Thai Quang Tung, Doheon Lee, Dae-Won Kim, Jong-Tae Lim</i>	180
Comparison of Tree Based Methods on Mammography Data <i>Richard De Veaux, Thu Hoàng</i>	186

Bayesian Sequence Learning for Predicting Protein Cleavage Points <i>Michael Mayo</i>	192
--	-----

A Novel Indexing Method for Efficient Sequence Matching in Large DNA Database Environment <i>Jung-Im Won, Jee-Hee Yoon, Sanghyun Park, Sang-Wook Kim</i>	203
---	-----

Classification and Ranking

Threshold Tuning for Improved Classification Association Rule Mining <i>Frans Coenen, Paul Leng, Lu Zhang</i>	216
--	-----

Using Rough Set in Feature Selection and Reduction in Face Recognition Problem <i>Le Hoai Bac, Nguyen Anh Tuan</i>	226
---	-----

Analysis of Company Growth Data Using Genetic Algorithms on Binary Trees <i>Gerrit K. Janssens, Kenneth Sørensen, Arthur Limère, Koen Vanhoof</i>	234
--	-----

Considering Re-occurring Features in Associative Classifiers <i>Rafal Rak, Wojciech Stach, Osmar R. Zaiane, Maria-Luiza Antonie</i>	240
--	-----

A New Evolutionary Neural Network Classifier <i>Arit Thammano, Asavin Meengen</i>	249
--	-----

A Privacy-Preserving Classification Mining Algorithm <i>Weiping Ge, Wei Wang, Xiaorong Li, Baile Shi</i>	256
---	-----

Combining Classifiers with Multi-representation of Context in Word Sense Disambiguation <i>Cuong Anh Le, Va-Nam Huynh, Akira Shimazu</i>	262
---	-----

Automatic Occupation Coding with Combination of Machine Learning and Hand-Crafted Rules <i>Kazuko Takahashi, Hiroya Takamura, Manabu Okumura</i>	269
---	-----

Retrieval Based on Language Model with Relative Entropy and Feedback <i>Hua Huo, Boqin Feng</i>	280
--	-----

Text Classification for DAG-Structured Categories <i>Cao D. Nguyen, Tran A. Dung, Tru H. Cao</i>	290
---	-----

Sentiment Classification Using Word Sub-sequences and Dependency Sub-trees <i>Shotaro Matsumoto, Hiroya Takamura, Manabu Okumura</i>	301
---	-----

Improving Rough Classifiers Using Concept Ontology
Nguyen Sinh Hoa, Nguyen Hung Son 312

QED: An Efficient Framework for Temporal Region Query Processing
Yi-Hong Chu, Kun-Ta Chuang, Ming-Syan Chen 323

Clustering

A MPAA-Based Iterative Clustering Algorithm Augmented by Nearest
 Neighbors Search for Time-Series Data Streams
*Jessica Lin, Michai Vlachos, Eamonn Keogh, Dimitrios Gunopulos,
 Jianwei Liu, Shoujian Yu, Jiajin Le* 333

Locating Motifs in Time-Series Data
Zheng Liu, Jeffrey Xu Yu, Xuemin Lin, Hongjun Lu, Wei Wang 343

Stochastic Local Clustering for Massive Graphs
Satu Elisa Schaeffer 354

A Neighborhood-Based Clustering Algorithm
Shuigeng Zhou, Yue Zhao, Jihong Guan, Joshua Huang 361

Improved Self-splitting Competitive Learning Algorithm
Jun Liu, Kotagiri Ramamohanarao 372

Speeding-Up Hierarchical Agglomerative Clustering in Presence
 of Expensive Metrics
Mirco Nanni 378

Dynamic Cluster Formation Using Level Set Methods
Andy M. Yip, Chris Ding, Tony F. Chan 388

A Vector Field Visualization Technique for Self-organizing Maps
Georg Pözlbauer, Andreas Rauber, Michael Dittenbach 399

Visualization of Cluster Changes by Comparing Self-organizing Maps
Denny McG. Scurie, David McG. Squire 410

An Incremental Data Stream Clustering Algorithm Based
 on Dense Units Detection
Jing Gao, Jianzhong Li, Zhaogong Zhang, Pang-Ning Tan 420

Visual Interactive Evolutionary Algorithm for High Dimensional Data
 Clustering and Outlier Detection
Lydia Boudjeloud, François Poulet 426

Approximated Clustering of Distributed High-Dimensional Data <i>Hans-Peter Kriegel, Peter Kunath, Martin Pfeifle, Matthias Renz</i>	432
--	-----

Dynamic Data Mining

Improvements of IncSpan: Incremental Mining of Sequential Patterns in Large Database <i>Son N. Nguyen, Xingzhi Sun, Maria E. Orłowska</i>	442
---	-----

Efficient Sampling: Application to Image Data <i>Surong Wang, Manoranjan Dash, Liang-Tien Chia</i>	452
---	-----

Cluster-Based Rough Set Construction <i>Qiang Li, Bo Zhang</i>	464
---	-----

Graphic Model Discovery

Learning Bayesian Networks Structures from Incomplete Data: An Efficient Approach Based on Extended Evolutionary Programming <i>Xiaolin Li, Xiangdong He, Senmiao Yuan</i>	474
--	-----

Dynamic Fuzzy Clustering for Recommender Systems <i>Sung-Hwan Min, Ingoo Han</i>	480
---	-----

Improving Mining Quality by Exploiting Data Dependency <i>Fang Chu, Yizhou Wang, Carlo Zaniolo, D. Stott Parker</i>	486
--	-----

High Dimensional Data

Feature Selection for High Dimensional Face Image Using Self-organizing Maps <i>Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, Fuyan Zhang</i>	500
---	-----

Progressive Sampling for Association Rules Based on Sampling Error Estimation <i>Kun-Ta Chuang, Ming-Syan Chen, Wen-Chieh Yang</i>	505
--	-----

<i>CLeVer</i> : A Feature Subset Selection Technique for Multivariate Time Series <i>Kiyoung Yang, Hyunjin Yoon, Cyrus Shahabi</i>	516
---	-----

Covariance and PCA for Categorical Variables <i>Hiroataka Niitsuma, Takashi Okada</i>	523
--	-----

Integration of Data Warehousing

ADenTS: An Adaptive Density-Based Tree Structure for Approximating Aggregate Queries over Real Attributes <i>Tianyi Wu, Jian Xu, Chen Wang, Wei Wang, Baile Shi</i>	529
Frequent Itemset Mining with Parallel RDBMS <i>Xuequn Shang, Kai-Uwe Sattler</i>	539

Knowledge Management

Using Consensus Susceptibility and Consistency Measures for Inconsistent Knowledge Management <i>Ngoc Thanh Nguyen, Michal Malowiecki</i>	545
WLPMiner: Weighted Frequent Pattern Mining with Length-Decreasing Support Constraints <i>Unil Yun, John J. Leggett</i>	555

Machine Learning Methods

A Framework for Incorporating Class Priors into Discriminative Classification <i>Rong Jin, Yi Liu</i>	568
Increasing Classification Accuracy by Combining Adaptive Sampling and Convex Pseudo-Data <i>Chia Huey Ooi, Madhu Chetty</i>	578
Kernels over Relational Algebra Structures <i>Adam Woźnica, Alexandros Kalousis, Melanie Hilario</i>	588
Adaptive Nonlinear Auto-Associative Modeling Through Manifold Learning <i>Junping Zhang, Stan Z. Li</i>	599
Maximizing Tree Diversity by Building Complete-Random Decision Trees <i>Fei Tony Liu, Kai Ming Ting, Wei Fan</i>	605
SETRED: Self-training with Editing <i>Ming Li, Zhi-Hua Zhou</i>	611
Adjusting Mixture Weights of Gaussian Mixture Model via Regularized Probabilistic Latent Semantic Analysis <i>Luo Si, Rong Jin</i>	622

Training Support Vector Machines Using Greedy Stagewise Algorithm <i>Liefeng Bo, Ling Wang, Licheng Jiao</i>	632
CI-GBI: A Novel Approach for Extracting Typical Patterns from Graph-Structured Data <i>Phu Chien Nguyen, Kouzou Ohara, Hiroshi Motoda, Takashi Washio</i>	639
Improved Bayesian Spam Filtering Based on Co-weighted Multi-area Information <i>Raju Shrestha, Yaping Lin</i>	650

Novel Algorithms

An Efficient Framework for Mining Flexible Constraints <i>Arnaud Soulet, Bruno Crémilleux</i>	661
Support Oriented Discovery of Generalized Disjunction-Free Representation of Frequent Patterns with Negation <i>Marzena Kryszkiewicz, Katarzyna Cichoń</i>	672
Feature Selection Algorithm for Data with Both Nominal and Continuous Features <i>Wenyin Tang, Kezhi Mao</i>	683
A Two-Phase Algorithm for Fast Discovery of High Utility Itemsets <i>Ying Liu, Wei-keng Liao, Alok Choudhary</i>	689
On Multiple Query Optimization in Data Mining <i>Marek Wojciechowski, Maciej Zakrzewicz</i>	696
USAID: Unifying Signature-Based and Anomaly-Based Intrusion Detection <i>Zhuowei Li, Amitabha Das, Jianying Zhou</i>	702

Spatial Data

Mining Mobile Group Patterns: A Trajectory-Based Approach <i>San-Yih Hwang, Ying-Han Liu, Jeng-Kuen Chiu, Ee-Peng Lim</i>	713
Can We Apply Projection Based Frequent Pattern Mining Paradigm to Spatial Co-location Mining? <i>Yan Huang, Liqin Zhang, Ping Yu</i>	719
PatZip: Pattern-Preserved Spatial Data Compression <i>Yu Qian, Kang Zhang, D. T. Huynh</i>	726

Temporal Data

A Likelihood Ratio Distance Measure for the Similarity Between the Fourier Transform of Time Series <i>Gareth J. Janacek, Anthony J. Bagnall, Michael Powell</i>	737
The TIMERS II Algorithm for the Discovery of Causality <i>Howard J. Hamilton, Kamran Karimi</i>	744
A Recent-Based Dimension Reduction Technique for Time Series Data <i>Yanchang Zhao, Chengqi Zhang, Shichao Zhang</i>	751
Graph Partition Model for Robust Temporal Data Segmentation <i>Jinhui Yuan, Bo Zhang, Fuzong Lin</i>	758
Accurate Symbolization of Time Series <i>Xinqiang Zuo, Xiaoming Jin</i>	764
A Novel Bit Level Time Series Representation with Implication of Similarity Search and Clustering <i>Chotirat Ratanamahatana, Eamonn Keogh, Anthony J. Bagnall, Stefano Lonardi</i>	771
Finding Temporal Features of Event-Oriented Patterns <i>Xingzhi Sun, Maria E. Orlowska, Xue Li</i>	778
An Anomaly Detection Method for Spacecraft Using Relevance Vector Learning <i>Ryohei Fujimaki, Takehisa Yairi, Kazuo Machida</i>	785
Cyclic Pattern Kernels Revisited <i>Tamás Horváth</i>	791

Text and Web Data Mining

Subspace Clustering of Text Documents with Feature Weighting <i>K-Means Algorithm</i> <i>Liping Jing, Michael K. Ng, Jun Xu, Joshua Zhexue Huang</i>	802
Using Term Clustering and Supervised Term Affinity Construction to Boost Text Classification <i>Chong Wang, Wenyuan Wang</i>	813
Technology Trends Analysis from the Internet Resources <i>Shin-ichi Kobayashi, Yasuyuki Shirai, Kazuo Hiyane, Fumihiko Kumeno, Hiroshi Inujima, Noriyoshi Yamauchi</i>	820

Dynamic Mining Hierarchical Topic from Web News Stream Data Using Divisive-Agglomerative Clustering Method <i>Jian-Wei Liu, Shou-Jian Yu, Jia-Jin Le</i>	826
Collecting Topic-Related Web Pages for Link Structure Analysis by Using a Potential Hub and Authority First Approach <i>Leuo-Hong Wang, Tong-Wen Lee</i>	832
A Top-Down Algorithm for Mining Web Access Patterns from Web Logs <i>Jian-Kui Guo, Bei-jun Ruan, Zun-ping Cheng, Fang-zhong Su, Ya-qin Wang, Xu-bin Deng, Shang Ning, Yang-Yong Zhu</i>	838
Kernel Principal Component Analysis for Content Based Image Retrieval <i>Guang-Ho Cha</i>	844
Mining Frequent Trees with Node-Inclusion Constraints <i>Atsuyoshi Nakamura, Mineichi Kudo</i>	850
Author Index	861