

Table of Contents

Invited Papers

Warehousing and Mining Massive RFID Data Sets <i>Jiawei Han, Hector Gonzalez, Xiaolei Li, Diego Klabjan</i>	1
Self-organising Map Techniques for Graph Data Applications to Clustering of XML Documents <i>A.C. Tsoi, M. Hagenbuchner, A. Sperduti</i>	19
Finding Time Series Discords Based on Haar Transform <i>Ada Wai-chee Fu, Oscar Tat-Wing Leung, Eamonn Keogh, Jessica Lin</i>	31
Learning with Local Drift Detection <i>João Gama, Gladys Castillo</i>	42

Association Rules

A Fast Algorithm for Maintenance of Association Rules in Incremental Databases <i>Xin Li, Zhi-Hong Deng, Shiwei Tang</i>	56
Extending OLAP with Fuzziness for Effective Mining of Fuzzy Multidimensional Weighted Association Rules <i>Mehmet Kaya, Reda Alhajj</i>	64
Incremental Maintenance of Association Rules Based on Multiple Previously Mined Results <i>Zhuohua Duan, Zixing Cai, Yan Lv</i>	72
Mining and Validation of Localized Frequent Web Access Patterns with Dynamic Tolerance <i>Olfa Nasraoui, Suchandra Goswami</i>	80
SA-IFIM: Incrementally Mining Frequent Itemsets in Update Distorted Databases <i>Jinlong Wang, Congfu Xu, Hongwei Dan, Yunhe Pan</i>	92

XII Table of Contents

Study of Positive and Negative Association Rules Based on Multi-confidence and Chi-Squared Test <i>Xiangjun Dong, Fengrong Sun, Xiqing Han, Ruilian Hou</i>	100
Efficiently Mining Maximal Frequent Mutually Associated Patterns <i>Zhongmei Zhou, Zhaohui Wu, Chunshan Wang, Yi Feng</i>	110
Efficiently Mining Mutually and Positively Correlated Patterns <i>Zhongmei Zhou, Zhaohui Wu, Chunshan Wang, Yi Feng</i>	118
Classification	
ComEnVprs: A Novel Approach for Inducing Decision Tree Classifiers <i>Shuqin Wang, Jinmao Wei, Junping You, Dayou Liu</i>	126
Towards a Rough Classification of Business Travelers <i>Rob Law, Thomas Bauer, Karin Weber, Tony Tse</i>	135
Feature Extraction Based on Optimal Discrimination Plane in ECG Signal Classification <i>Dingfei Ge, Xiao Qu</i>	143
Music Style Classification with a Novel Bayesian Model <i>Yatong Zhou, Taiyi Zhang, Jiancheng Sun</i>	150
Classification of Polarimetric SAR Data Based on Multidimensional Watershed Clustering <i>Wen Yang, Hao Wang, Yongfeng Cao, Haijian Zhang</i>	157
An Effective Combination Based on Class-Wise Expertise of Diverse Classifiers for Predictive Toxicology Data Mining <i>Daniel Neagu, Gongde Guo, Shanshan Wang</i>	165
Robust Collective Classification with Contextual Dependency Network Models <i>Yonghong Tian, Tiejun Huang, Wen Gao</i>	173
User-Centered Image Semantics Classification <i>Hongli Xu, De Xu, Fangshi Wang</i>	181
A Performance Study of Gaussian Kernel Classifiers for Data Mining Applications <i>Miyoung Shin</i>	189

TTLSC – Transductive Total Least Square Model for Classification and Its Application in Medicine <i>Qun Song, Tian Min Ma, Nikola Kasabov</i>	197
Forecasting Electricity Market Price Spikes Based on Bayesian Expert with Support Vector Machines <i>Wei Wu, Jianzhong Zhou, Li Mo, Chenjun Zhu</i>	205
Integrating Local One-Class Classifiers for Image Retrieval <i>Yiqing Tu, Gang Li, Honghua Dai</i>	213
Incremental Discretization for Naïve-Bayes Classifier <i>Jingli Lu, Ying Yang, Geoffrey I. Webb</i>	223
Distance Guided Classification with Gene Expression Programming <i>Lei Duan, Changjie Tang, Tianqing Zhang, Dagang Wei, Huan Zhang</i>	239
Research on Multi-valued and Multi-labeled Decision Trees <i>Hong Li, Rui Zhao, Jianer Chen, Yao Xiang</i>	247
Clustering	
A Spatial Clustering Algorithm Based on SOFM <i>Zhong Qu, Lian Wang</i>	255
Mining Spatial-temporal Clusters from Geo-databases <i>Min Wang, Aiping Wang, Anbo Li</i>	263
A Fuzzy Subspace Algorithm for Clustering High Dimensional Data <i>Guojun Gan, Jianhong Wu, Zijiang Yang</i>	271
Robust Music Information Retrieval on Mobile Network Based on Multi-Feature Clustering <i>Won-Jung Yoon, Sanghun Oh, Kyu-Sik Park</i>	279
Joint Cluster Based Co-clustering for Clustering Ensembles <i>Tianming Hu, Liping Liu, Chao Qu, Sam Yuan Sung</i>	284
Mining Gait Pattern for Clinical Locomotion Diagnosis Based on Clustering Techniques <i>Guandong Xu, Yanchun Zhang, Rezaul Begg</i>	296
Combining Multiple Clusterings Via k-Modes Algorithm <i>Huilan Luo, Fansheng Kong, Yixiao Li</i>	308

XIV Table of Contents

HOV ³ : An Approach to Visual Cluster Analysis <i>Ke-Bing Zhang, Mehmet A. Orgun, Kang Zhang</i>	316
A New Fuzzy Co-clustering Algorithm for Categorization of Datasets with Overlapping Clusters <i>William-Chandra Tjhi, Lihui Chen</i>	328
Quantum-Behaved Particle Swarm Optimization Clustering Algorithm <i>Jun Sun, Wenbo Xu, Bin Ye</i>	340
Clustering Mixed Data Based on Evidence Accumulation <i>Huilan Luo, Fansheng Kong, Yixiao Li</i>	348
Mining Maximal Local Conserved Gene Clusters from Microarray Data <i>Yuhai Zhao, Guoren Wang, Ying Yin, Guangyu Xu</i>	356
Novel Algorithms	
A Novel P2P Information Clustering and Retrieval Mechanism <i>Huaxiang Zhang, Peide Liu</i>	364
Keeping Track of Customer Life Cycle to Build Customer Relationship <i>Sung Ho Ha, Sung Min Bae</i>	372
Mining of Flexible Manufacturing System Using Work Event Logs and Petri Nets <i>Hesuan Hu, Zhiwu Li, Anrong Wang</i>	380
Improved Genetic Algorithm for Multiple Sequence Alignment Using Segment Profiles (GASP) <i>Yanping Lv, Shaozi Li, Changle Zhou, Wenzhong Guo, Zhengming Xu</i>	388
A Novel Visual Clustering Algorithm for Finding Community in Complex Network <i>Shuzhong Yang, Siwei Luo, Jianyu Li</i>	396
Self-Organizing Network Evolving Model for Mining Network Community Structure <i>Bo Yang</i>	404
An Interactive Visualization Environment for Data Exploration Using Points of Interest <i>David Da Costa, Gilles Venturini</i>	416

Forecasting the Volatility of Stock Price Index <i>Tae Hyup Roh</i>	424
ExMiner: An Efficient Algorithm for Mining Top-K Frequent Patterns <i>Tran Minh Quang, Shigeru Oyanagi, Katsuhiko Yamazaki</i>	436
Learning Bayesian Networks Structure with Continuous Variables <i>Shuang-Cheng Wang, Xiao-Lin Li, Hai-Yan Tang</i>	448
A Unified Strategy of Feature Selection <i>Peng Liu, Naijun Wu, Jiaxian Zhu, Junjie Yin, Wei Zhang</i>	457
Experimental Comparison of Feature Subset Selection Using GA and ACO Algorithm <i>Keunjoon Lee, Jinu Joo, Jihoon Yang, Vasant Honavar</i>	465
OMVD: An Optimization of MVD <i>Zhi He, Shengfeng Tian, Houkuan Huang</i>	473
ZED: Explaining Temporal Variations in Query Volume <i>Maojin Jiang, Shlomo Argamon, Abdur Chowdhury, Kush Sidhu</i>	485
An Effective Multi-level Algorithm for Bisecting Graph <i>Ming Leng, Songnian Yu</i>	493
A New Polynomial Time Algorithm for Bayesian Network Structure Learning <i>Sanghack Lee, Jihoon Yang, Sungyong Park</i>	501
Personalized Recommendation Based on Partial Similarity of Interests <i>Ming-Hua Yang, Zhi-Min Gu</i>	509
A Fast Implementation of the EM Algorithm for Mixture of Multinomials <i>Jan Peter Patist</i>	517
A Novel Approach to Pattern Recognition Based on PCA-ANN in Spectroscopy <i>Xiaoli Li, Yong He</i>	525
Semi-supervised Dynamic Counter Propagation Network <i>Yao Chen, Yuntao Qian</i>	533
The Practical Method of Fractal Dimensionality Reduction Based on Z-Ordering Technique <i>Guanghui Yan, Zhanhuai Li, Liu Yuan</i>	542

Feature Selection for Complex Patterns <i>Peter Schenkel, Wanqing Li, Wanquan Liu</i>	550
Naïve Bayesian Tree Pruning by Local Accuracy Estimation <i>Zhipeng Xie</i>	558
A New Visualization Method for Patent Map: Application to Ubiquitous Computing Technology <i>Jong Hwan Suh, Sang Chan Park</i>	566
Local Linear Logistic Discriminant Analysis with Partial Least Square Components <i>Jangsun Baek, Young Sook Son</i>	574
Activity Mining: Challenges and Prospects <i>Longbing Cao</i>	582
Finding the Optimal Cardinality Value for Information Bottleneck Method <i>Gang Li, Dong Liu, Yiqing Tu, Yangdong Ye</i>	594
A New Algorithm for Enumerating All Maximal Cliques in Complex Network <i>Li Wan, Bin Wu, Nan Du, Qi Ye, Ping Chen</i>	606
Modeling and Mining the Rule Evolution <i>Ding Pan</i>	618
Knowledge Reduction in Inconsistent Decision Tables <i>Qihe Liu, Leitong Chen, Jianzhong Zhang, Fan Min</i>	626
Text Mining	
Semantic Scoring Based on Small-World Phenomenon for Feature Selection in Text Mining <i>Chong Huang, Yonghong Tian, Tiejun Huang, Wen Gao</i>	636
A Comparative Study on Text Clustering Methods <i>Yan Zheng, Xiaochun Cheng, Ronghuai Huang, Yi Man</i>	644
Concept Based Text Classification Using Labeled and Unlabeled Data <i>Ping Gu, Qingsheng Zhu, Xiping He</i>	652
Learning Semantic User Profiles from Text <i>M. DeGemmis, P. Lops, G. Semeraro</i>	661

Multimedia Mining

- Audiovisual Integration for Racquet Sports Video Retrieval
Yaqin Zhao, Xianzhong Zhou, Guizhong Tang 673
- A Correlation Approach for Automatic Image Annotation
*David R. Hardoon, Craig Saunders, Sandor Szedmak,
 John Shawe-Taylor* 681

Sequential Data Mining and Time Series Mining

- Fast Discovery of Time-Constrained Sequential Patterns Using
 Time-Indexes
Ming-Yen Lin, Sue-Chen Hsueh, Chia-Wen Chang 693
- Multi-dimensional Sequential Pattern Mining Based on Concept Lattice
Yang Jin, Wanli Zuo 702
- Mining Time-Delayed Coherent Patterns in Time Series Gene
 Expression Data
Linjun Yin, Guoren Wang, Keming Mao, Yuhai Zhao 711
- Mining Delay in Streaming Time Series of Industrial Process
Haijie Gu, Gang Rong 723
- Segmental Semi-Markov Model Based Online Series Pattern
 Detection Under Arbitrary Time Scaling
Guangjie Ling, Yuntao Qian, Sen Jia 731
- Diagnosis of Inverter Faults in PMSM DTC Drive Using
 Time-Series Data Mining Technique
Dan Sun, Jun Meng, Zongyuan He 741
- Applications of Data Mining Time Series to Power Systems
 Disturbance Analysis
Jun Meng, Dan Sun, Zhiyong Li 749
- Mining Compressed Sequential Patterns
Lei Chang, Dongqing Yang, Shiwei Tang, Tengjiao Wang 761
- Effective Feature Preprocessing for Time Series Forecasting
Jun Hua Zhao, ZhaoYang Dong, Zhao Xu 769
- On Similarity of Financial Data Series Based on Fractal Dimension
Jian-rong Hou, Hui Zhao, Pei Huang 782

Web Mining

A Hierarchical Model of Web Graph <i>Jie Han, Yong Yu, Chenxi Lin, Dingyi Han, Gui-Rong Xue</i>	790
Web Scale Competitor Discovery Using Mutual Information <i>Rui Li, Shenghua Bao, Jin Wang, Yuanjie Liu, Yong Yu</i>	798

Biomedical Mining

Co-expression Gene Discovery from Microarray for Integrative Systems Biology <i>Yutao Ma, Yonghong Peng</i>	809
Cardiovascular Disease Diagnosis Method by Emerging Patterns <i>Heon Gyu Lee, Kiyong Noh, Bum Ju Lee, Ho-Sun Shon, Keun Ho Ryu</i>	819
DNA Microarray Data Clustering by Hidden Markov Models and Bayesian Information Criterion <i>Phasit Charoenkwan, Aompilai Manorat, Jeerayut Chaijaruwanich, Sukon Prasitwattanaseree, Sakarindr Bhumiratana</i>	827
Application of Factor Analysis on <i>Mycobacterium Tuberculosis</i> Transcriptional Responses for Drug Clustering, Drug Target and Pathway Detections <i>Jeerayut Chaijaruwanich, Jamlong Khamphachua, Sukon Prasitwattanaseree, Saradee Warit, Prasit Palittapongarnpim</i>	835
First Steps to an Audio Ontology-Based Classifier for Telemedicine <i>Cong Phuong Nguyen, Ngoc Yen Pham, Eric Castelli</i>	845
Obstacles and Misunderstandings Facing Medical Data Mining <i>Ashkan Sami</i>	856
SVM-Based Tumor Classification with Gene Expression Data <i>Shulin Wang, Ji Wang, Huowang Chen, Boyun Zhang</i>	864
GEPCLASS: A Classification Rule Discovery Tool Using Gene Expression Programming <i>Wagner R. Weinert, Heitor S. Lopes</i>	871

Advanced Applications

CBR-Based Knowledge Discovery on Results of Evolutionary Design of Logic Circuits <i>Shuguang Zhao, Mingying Zhao, Jin Li, Change Wang</i>	881
Data Summarization Approach to Relational Domain Learning Based on Frequent Pattern to Support the Development of Decision Making <i>Rayner Alfred, Dimitar Kazakov</i>	889
Extreme Value Dependence in Problems with a Changing Causation Structure <i>Marlon Núñez, Rafael Morales</i>	899
A Study on Object Recognition Technology Using PCA in the Variable Illumination <i>Jong-Min Kim, Hwan-Seok Yang</i>	911
Pattern Recurring in Three-Dimensional Graph Based on Data Mining <i>Yanbing Liu, Menghao Wang</i>	919
Mining the Useful Skyline Set Based on the Acceptable Difference <i>Zhenhua Huang, Wei Wang</i>	927
Modeling Information-Sharing Behaviors in BitTorrent System Based on Real Measurement <i>Jinkang Jia, Changjia Chen</i>	934
Financial Distress Prediction Based on Similarity Weighted Voting CBR <i>Jie Sun, Xiao-Feng Hui</i>	947
Customer Churn Prediction by Hybrid Model <i>Jae Sik Lee, Jin Chun Lee</i>	959
Base Vector Selection for Kernel Matching Pursuit <i>Qing Li, Licheng Jiao</i>	967
<i>WaveSim</i> Transform for Multi-channel Signal Data Mining Through Linear Regression PCA <i>R. Pradeep Kumar, P. Nagabhushan</i>	977
Research on Query-by-Committee Method of Active Learning and Application <i>Yue Zhao, Ciwen Xu, Yongcun Cao</i>	985

Traffic Management Genetic Algorithm Supporting Data Mining and QoS in Sensor Networks
Yantao Pan, Wei Peng, Xicheng Lu 992

Comparison of Data Pre-processing in Pattern Recognition of Milk Powder Vis/NIR Spectra
Haiyan Cen, Yidan Bao, Min Huang, Yong He 1000

Semi-automatic Hot Event Detection
Tingting He, Guozhong Qu, Siwei Li, Xinhui Tu, Yong Zhang, Han Ren 1008

Security and Privacy Issues

Profile-Based Security Against Malicious Mobile Agents
Hua Li, Glenna Greene, Rafael Alonso 1017

A Comprehensive Categorization of DDoS Attack and DDoS Defense Techniques
Usman Tariq, ManPyo Hong, Kyung-suk Lhee 1025

Structural Analysis and Mathematical Methods for Destabilizing Terrorist Networks Using Investigative Data Mining
Nasrullah Memon, Henrik Legind Larsen 1037

Alert Correlation Analysis in Intrusion Detection
Moon Sun Shin, Kyeong Ja Jeong 1049

Spatial Data Mining

OSDM: Optimized Shape Distribution Method
Ashkan Sami, Ryoichi Nagatomi, Makoto Takahashi, Takeshi Tokuyama 1057

View-Angle of Spatial Data Mining
Shuliang Wang, Haning Yuan 1065

Streaming Data Mining

Maintaining Moving Sums over Data Streams
Tzu-Chiang Wu, Arbee L.P. Chen 1077

MFIS—Mining Frequent Itemsets on Data Streams
Zhi-jun Xie, Hong Chen, Cuiping Li 1085

Improving the Performance of Data Stream Classifiers by Mining
Recurring Contexts
 Yong Wang, Zhanhuai Li, Yang Zhang, Longbo Zhang, Yun Jiang 1094

Author Index 1107