

# Table of Contents

## Keynote Speech

Graph Mining: Laws, Generators and Tools .....	1
<i>Christos Faloutsos</i>	

## Invited Speeches

Efficient Algorithms for Mining Frequent and Closed Patterns from Semi-structured Data.....	2
<i>Hiroki Arimura</i>	
Supporting Creativity: Towards Associative Discovery of New Insights.....	14
<i>Michael R. Berthold, Fabian Dill, Tobias Kötter, and Kilian Thiel</i>	
Cost-Sensitive Classifier Evaluation Using Cost Curves .....	26
<i>Robert C. Holte and Chris Drummond</i>	
Prospective Scientific Methodology in Knowledge Society .....	30
<i>Genshiro Kitagawa</i>	

## Long Papers

SubClass: Classification of Multidimensional Noisy Data Using Subspace Clusters.....	40
<i>Ira Assent, Ralph Krieger, Petra Welter, Jörg Herbers, and Thomas Seidl</i>	
Mining Quality-Aware Subspace Clusters .....	53
<i>Ying-Ju Chen, Yi-Hong Chu, and Ming-Syan Chen</i>	
A Decremental Approach for Mining Frequent Itemsets from Uncertain Data .....	64
<i>Chun-Kit Chui and Ben Kao</i>	
Multi-class Named Entity Recognition Via Bootstrapping with Dependency Tree-Based Patterns .....	76
<i>Van B. Dang and Akiko Aizawa</i>	
Towards Region Discovery in Spatial Datasets .....	88
<i>Wei Ding, Rachsuda Jiamthaphaksin, Rachana Parmar, Dan Jiang, Tomasz F. Stepinski, and Christoph F. Eick</i>	

Accurate and Efficient Retrieval of Multimedia Time Series Data Under Uniform Scaling and Time Warping .....	100
<i>Waiyawuth Euachongprasit and Chotirat Ann Ratanamahatana</i>	
Feature Construction Based on Closedness Properties Is Not That Simple.....	112
<i>Dominique Gay, Nazha Selmaoui, and Jean-François Boulicaut</i>	
On Addressing Accuracy Concerns in Privacy Preserving Association Rule Mining.....	124
<i>Ling Guo, Songtao Guo, and Xintao Wu</i>	
Privacy-Preserving Linear Fisher Discriminant Analysis.....	136
<i>Shuguo Han and Wee Keong Ng</i>	
Unsupervised Change Analysis Using Supervised Learning .....	148
<i>Shohei Hido, Tsuyoshi Idé, Hisashi Kashima, Harunobu Kubo, and Hirofumi Matsuzawa</i>	
ANEMI: An Adaptive Neighborhood Expectation-Maximization Algorithm with Spatial Augmented Initialization .....	160
<i>Tianming Hu, Hui Xiong, Xueqing Gong, and Sam Yuan Sung</i>	
Minimum Variance Associations — Discovering Relationships in Numerical Data.....	172
<i>Szymon Jaroszewicz</i>	
An Efficient Unordered Tree Kernel and Its Application to Glycan Classification .....	184
<i>Tetsuji Kuboyama, Kouichi Hirata, and Kiyoko F. Aoki-Kinoshita</i>	
Generation of Globally Relevant Continuous Features for Classification .....	196
<i>Sylvain Létourneau, Stan Matwin, and A. Fazel Famili</i>	
Mining Bulletin Board Systems Using Community Generation .....	209
<i>Ming Li, Zhongfei (Mark) Zhang, and Zhi-Hua Zhou</i>	
Extreme Support Vector Machine Classifier .....	222
<i>Qiuge Liu, Qing He, and Zhongzhi Shi</i>	
LCM over ZBDDs: Fast Generation of Very Large-Scale Frequent Itemsets Using a Compact Graph-Based Representation .....	234
<i>Shin-ichi Minato, Takeaki Uno, and Hiroki Arimura</i>	
Unusual Pattern Detection in High Dimensions .....	247
<i>Minh Quoc Nguyen, Leo Mark, and Edward Omiecinski</i>	
Person Name Disambiguation in Web Pages Using Social Network, Compound Words and Latent Topics .....	260
<i>Shingo Ono, Issei Sato, Minoru Yoshida, and Hiroshi Nakagawa</i>	

Mining Correlated Subgraphs in Graph Databases .....	272
<i>Tomonobu Ozaki and Takenao Ohkawa</i>	
A Minimal Description Length Scheme for Polynomial Regression .....	284
<i>Aleksandar Pečkov, Sašo Džeroski, and Ljupčo Todorovski</i>	
Handling Numeric Attributes in Hoeffding Trees .....	296
<i>Bernhard Pfahringer, Geoffrey Holmes, and Richard Kirkby</i>	
Scaling Record Linkage to Non-uniform Distributed Class Sizes .....	308
<i>Steffen Rendle and Lars Schmidt-Thieme</i>	
Large-Scale k-Means Clustering with User-Centric Privacy Preservation .....	320
<i>Jun Sakuma and Shigenobu Kobayashi</i>	
Semi-Supervised Local Fisher Discriminant Analysis for Dimensionality Reduction .....	333
<i>Masashi Sugiyama, Tsuyoshi Idé, Shinichi Nakajima, and Jun Sese</i>	
An Efficient Algorithm for Finding Similar Short Substrings from Large Scale String Data .....	345
<i>Takeaki Uno</i>	
Ambiguous Frequent Itemset Mining and Polynomial Delay Enumeration .....	357
<i>Takeaki Uno and Hiroki Arimura</i>	
Characteristic-Based Descriptors for Motion Sequence Recognition .....	369
<i>Liang Wang, Xiaozhe Wang, Christopher Leckie, and Kotagiri Ramamohanarao</i>	
Protecting Privacy in Incremental Maintenance for Distributed Association Rule Mining .....	381
<i>W.K. Wong, David W. Cheung, Edward Hung, and Huan Liu</i>	
SEM: Mining Spatial Events from the Web .....	393
<i>Kaifeng Xu, Rui Li, Shenghua Bao, Dingyi Han, and Yong Yu</i>	
BOAI: Fast Alternating Decision Tree Induction Based on Bottom-Up Evaluation .....	405
<i>Bishan Yang, Tengjiao Wang, Dongqing Yang, and Lei Chang</i>	
Feature Selection by Nonparametric Bayes Error Minimization .....	417
<i>Shuang-Hong Yang and Bao-Gang Hu</i>	
A Framework for Modeling Positive Class Expansion with Single Snapshot .....	429
<i>Yang Yu and Zhi-Hua Zhou</i>	

## XVIII Table of Contents

A Decomposition Algorithm for Learning Bayesian Network Structures from Data .....	441
<i>Yifeng Zeng and Jorge Cordero Hernandez</i>	
Learning Classification Rules for Multiple Target Attributes .....	454
<i>Bernard Ženko and Sašo Džeroski</i>	
A Mixture Model for Expert Finding .....	466
<i>Jing Zhang, Jie Tang, Liu Liu, and Juanzi Li</i>	
On Privacy in Time Series Data Mining .....	479
<i>Ye Zhu, Yongjian Fu, and Huirong Fu</i>	

## Regular Papers

Exploiting Propositionalization Based on Random Relational Rules for Semi-supervised Learning .....	494
<i>Grant Anderson and Bernhard Pfahringer</i>	
On Discrete Data Clustering .....	503
<i>Nizar Bouguila and Walid ElGuebaly</i>	
Automatic Training Example Selection for Scalable Unsupervised Record Linkage .....	511
<i>Peter Christen</i>	
Analyzing PETs on Imbalanced Datasets When Training and Testing Class Distributions Differ .....	519
<i>David Cieslak and Nitesh Chawla</i>	
Improving the Robustness to Outliers of Mixtures of Probabilistic PCAs .....	527
<i>Nicolas Delannay, Cédric Archambeau, and Michel Verleysen</i>	
Exploratory Hot Spot Profile Analysis Using Interactive Visual Drill-Down Self-Organizing Maps .....	536
<i>Denny, Graham J. Williams, and Peter Christen</i>	
Maintaining Optimal Multi-way Splits for Numerical Attributes in Data Streams .....	544
<i>Tapio Elomaa and Petri Lehtinen</i>	
Efficient Mining of High Utility Itemsets from Large Datasets .....	554
<i>Alva Erwin, Raj P. Gopalan, and N.R. Achuthan</i>	
Tradeoff Analysis of Different Markov Blanket Local Learning Approaches .....	562
<i>Shunkai Fu and Michel C. Desmarais</i>	

Forecasting Urban Air Pollution Using HMM-Fuzzy Model .....	572
<i>M. Maruf Hossain, Md. Raful Hassan, and Michael Kirley</i>	
Relational Pattern Mining Based on Equivalent Classes of Properties Extracted from Samples .....	582
<i>Nobuhiro Inuzuka, Jun-ichi Motoyama, Shinpei Urazawa, and Tomofumi Nakano</i>	
Evaluating Standard Techniques for Implicit Diversity .....	592
<i>Ulf Johansson, Tuve Löfström, and Lars Niklasson</i>	
A Simple Characterization on Serially Constructible Episodes .....	600
<i>Takashi Kato and Kouichi Hirata</i>	
Bootstrap Based Pattern Selection for Support Vector Regression .....	608
<i>Dongil Kim and Sungzoon Cho</i>	
Tracking Topic Evolution in On-Line Postings: 2006 IBM Innovation Jam Data .....	616
<i>Mei Kobayashi and Raylene Yung</i>	
PAID: Packet Analysis for Anomaly Intrusion Detection .....	626
<i>Kuo-Chen Lee, Jason Chang, and Ming-Syan Chen</i>	
A Comparison of Different Off-Centered Entropies to Deal with Class Imbalance for Decision Trees .....	634
<i>Philippe Lenca, Stéphane Lallich, Thanh-Nghi Do, and Nguyen-Khang Pham</i>	
FISViz: A Frequent Itemset Visualizer .....	644
<i>Carson Kai-Sang Leung, Pourang P. Irani, and Christopher L. Carmichael</i>	
A Tree-Based Approach for Frequent Pattern Mining from Uncertain Data .....	653
<i>Carson Kai-Sang Leung, Mark Anthony F. Mateo, and Dale A. Brajczuk</i>	
Connectivity Based Stream Clustering Using Localised Density Exemplars .....	662
<i>Sebastian Lühr and Mihai Lazarescu</i>	
Learning User Purchase Intent from User-Centric Data .....	673
<i>Rajan Lukose, Jiye Li, Jing Zhou, and Satyanarayana Raju Penmetsa</i>	
Query Expansion for the Language Modelling Framework Using the Naïve Bayes Assumption .....	681
<i>Laurence A.F. Park and Kotagiri Ramamohanarao</i>	

Fast Online Estimation of the Joint Probability Distribution .....	689
<i>J.P. Patist</i>	
Fast $k$ Most Similar Neighbor Classifier for Mixed Data Based on Approximating and Eliminating.....	697
<i>Selene Hernández-Rodríguez, J. Ariel Carrasco-Ochoa, and J.Fco. Martínez-Trinidad</i>	
Entity Network Prediction Using Multitype Topic Models .....	705
<i>Hitohiro Shiozaki, Koji Eguchi, and Takenao Ohkawa</i>	
Using Supervised and Unsupervised Techniques to Determine Groups of Patients with Different Doctor-Patient Stability .....	715
<i>Eu-Gene Siew, Leonid Churilov, Kate A. Smith-Miles, and Joachim P. Sturmberg</i>	
Local Projection in Jumping Emerging Patterns Discovery in Transaction Databases .....	723
<i>Pawel Terlecki and Krzysztof Walczak</i>	
Applying Latent Semantic Indexing in Frequent Itemset Mining for Document Relation Discovery .....	731
<i>Thanaruk Theeramunkong, Kritsada Sripaew, and Manabu Okumura</i>	
G-TREACLE: A New Grid-Based and Tree-Alike Pattern Clustering Technique for Large Databases .....	739
<i>Cheng-Fa Tsai and Chia-Chen Yen</i>	
A Clustering-Oriented Star Coordinate Translation Method for Reliable Clustering Parameterization .....	749
<i>Chieh-Yuan Tsai and Chuang-Cheng Chiu</i>	
Constrained Clustering for Gene Expression Data Mining .....	759
<i>Vincent S. Tseng, Lien-Chin Chen, and Ching-Pin Kao</i>	
Concept Lattice-Based Mutation Control for Reactive Motifs Discovery .....	767
<i>Kitsana Waiyamai, Peera Liewlom, Thanapat Kangkachit, and Thanawin Rakthanmanon</i>	
Mining a Complete Set of Both Positive and Negative Association Rules from Large Databases .....	777
<i>Hao Wang, Xing Zhang, and Guoqing Chen</i>	
Designing a System for a Process Parameter Determined through Modified PSO and Fuzzy Neural Network .....	785
<i>Jui-Tsung Wong, Kuei-Hsien Chen, and Chwen-Tzeng Su</i>	

Data-Aware Clustering Hierarchy for Wireless Sensor Networks .....	795
<i>Xiaochen Wu, Peng Wang, Wei Wang, and Baile Shi</i>	
A More Topologically Stable Locally Linear Embedding Algorithm Based on R*-Tree .....	803
<i>Tian Xia, Jintao Li, Yongdong Zhang, and Sheng Tang</i>	
Sparse Kernel-Based Feature Weighting .....	813
<i>Shuang-Hong Yang, Yu-Jiu Yang, and Bao-Gang Hu</i>	
Term Committee Based Event Identification within News Topics .....	821
<i>Kuo Zhang, JuanZi Li, Gang Wu, and KeHong Wang</i>	
Locally Linear Online Mapping for Mining Low-Dimensional Data Manifolds .....	830
<i>Huicheng Zheng, Wei Shen, Qionghai Dai, and Sanqing Hu</i>	
A Creditable Subspace Labeling Method Based on D-S Evidence Theory .....	839
<i>Yu Zong, Xian-Chao Zhang, He Jiang, and Ming-Chu Li</i>	

## Short Papers

Discovering New Orders of the Chemical Elements through Genetic Algorithms .....	849
<i>Alexandre Blansch� and Shuichi Iwata</i>	
What Is Frequent in a Single Graph? .....	858
<i>Bj�rn Bringmann and Siegfried Nijssen</i>	
A Cluster-Based Genetic-Fuzzy Mining Approach for Items with Multiple Minimum Supports .....	864
<i>Chun-Hao Chen, Tzung-Pei Hong, and Vincent S. Tseng</i>	
A Selective Classifier for Incomplete Data .....	870
<i>Jingnian Chen, Houkuan Huang, Fengzhan Tian, and Shengfeng Tian</i>	
Detecting Near-Duplicates in Large-Scale Short Text Databases .....	877
<i>Caichun Gong, Yulan Huang, Xueqi Cheng, and Shuo Bai</i>	
Customer Churn Time Prediction in Mobile Telecommunication Industry Using Ordinal Regression .....	884
<i>Rupesh K. Gopal and Saroj K. Meher</i>	
Rule Extraction with Rough-Fuzzy Hybridization Method .....	890
<i>Nan-Chen Hsieh</i>	
I/O Scalable Bregman Co-clustering .....	896
<i>Kuo-Wei Hsu, Arindam Banerjee, and Jaideep Srivastava</i>	

## XXII Table of Contents

Jumping Emerging Patterns with Occurrence Count in Image Classification .....	904
<i>Lukasz Kobyliński and Krzysztof Walczak</i>	
Mining Non-coincidental Rules without a User Defined Support Threshold .....	910
<i>Yun Sing Koh</i>	
Transaction Clustering Using a Seeds Based Approach .....	916
<i>Yun Sing Koh and Russel Pears</i>	
Using Ontology-Based User Preferences to Aggregate Rank Lists in Web Search .....	923
<i>Lin Li, Zhenglu Yang, and Masaru Kitsuregawa</i>	
The Application of Echo State Network in Stock Data Mining .....	932
<i>Xiaowei Lin, Zehong Yang, and Yixu Song</i>	
Text Categorization of Multilingual Web Pages in Specific Domain .....	938
<i>Jicheng Liu and Chunyan Liang</i>	
Efficient Joint Clustering Algorithms in Optimization and Geography Domains .....	945
<i>Chia-Hao Lo and Wen-Chih Peng</i>	
Active Learning with Misclassification Sampling Using Diverse Ensembles Enhanced by Unlabeled Instances .....	951
<i>Jun Long, Jianping Yin, En Zhu, and Wentao Zhao</i>	
A New Model for Image Annotation .....	958
<i>Sanparith Marukatat</i>	
Unmixed Spectrum Clustering for Template Composition in Lung Sound Classification .....	964
<i>Tomonari Masada, Senya Kiyasu, and Sueharu Miyahara</i>	
Forward Semi-supervised Feature Selection .....	970
<i>Jiangtao Ren, Zhengyuan Qiu, Wei Fan, Hong Cheng, and Philip S. Yu</i>	
Automatic Extraction of Basis Expressions That Indicate Economic Trends .....	977
<i>Hiroki Sakaji, Hiroyuki Sakai, and Shigeru Masuyama</i>	
A New Framework for Taxonomy Discovery from Text .....	985
<i>Ahmad El Sayed, Hakim Hacid, and Djamel Zighed</i>	
R-Map: Mapping Categorical Data for Clustering and Visualization Based on Reference Sets .....	992
<i>Zhi-Yong Shen, Jun Sun, Yi-Dong Shen, and Ming Li</i>	



Mining Changes in Patent Trends for Competitive Intelligence .....	999
<i>Meng-Jung Shih, Duen-Ren Liu, and Ming-Li Hsu</i>	
Seeing Several Stars: A Rating Inference Task for a Document Containing Several Evaluation Criteria .....	1006
<i>Kazutaka Shimada and Tsutomu Endo</i>	
Structure-Based Hierarchical Transformations for Interactive Visual Exploration of Social Networks .....	1015
<i>Lisa Singh, Mitchell Beard, Brian Gopalan, and Gregory Nelson</i>	
CP-Tree: A Tree Structure for Single-Pass Frequent Pattern Mining ....	1022
<i>Syed Khairuzzaman Tanbeer, Chowdhury Farhan Ahmed, Byeong-Soo Jeong, and Young-Koo Lee</i>	
Combining Context and Existing Knowledge When Recognizing Biological Entities – Early Results .....	1028
<i>Mika Timonen and Antti Pesonen</i>	
Semantic Video Annotation by Mining Association Patterns from Visual and Speech Features .....	1035
<i>Vincent S. Tseng, Ja-Hwung Su, Jhih-Hong Huang, and Chih-Jen Chen</i>	
Cell-Based Outlier Detection Algorithm: A Fast Outlier Detection Algorithm for Large Datasets .....	1042
<i>You Wan and Fuling Bian</i>	
Fighting WebSpam: Detecting Spam on the Graph Via Content and Link Features .....	1049
<i>Yu-Jiu Yang, Shuang-Hong Yang, and Bao-Gang Hu</i>	
A Framework for Discovering Spatio-temporal Cohesive Networks .....	1056
<i>Jin Soung Yoo and Joengmin Hwang</i>	
Efficient Mining of Minimal Distinguishing Subgraph Patterns from Graph Databases .....	1062
<i>Zhiping Zeng, Jianyong Wang, and Lizhu Zhou</i>	
Combined Association Rule Mining .....	1069
<i>Huafeng Zhang, Yanchang Zhao, Longbing Cao, and Chengqi Zhang</i>	
Enriching WordNet with Folksonomies .....	1075
<i>Hao Zheng, Xian Wu, and Yong Yu</i>	
A New Credit Scoring Method Based on Rough Sets and Decision Tree .....	1081
<i>XiYue Zhou, DeFu Zhang, and Yi Jiang</i>	

XXIV Table of Contents

Analyzing the Propagation of Influence and Concept Evolution in  
Enterprise Social Networks through Centrality and Latent Semantic  
Analysis ..... 1090  
    *Weizhong Zhu, Chaomei Chen, and Robert B. Allen*

**Author Index** ..... 1099