

# **2006 IEEE International Conference on Data Mining (ICDM)**

**Hong Kong, China  
18 – 22 December 2006**

**Pages 1-627**



**IEEE Catalog Number: CFP06278-PRT  
ISBN: 978-1-4244-3077-2**

# TABLE OF CONTENTS

## **ICDM' 06: INVITED PAPERS**

<b>Neuroscience: New Insights for AI?</b> .....	1
<i>Tomaso Poggio</i>	
<b>Exploratory Mining in Cube Space</b> .....	4
<i>Raghu Ramakrishnan</i>	
<b>Data Mining Methods for Modeling Gene Expression Regulation and their Applications</b> .....	5
<i>Weixiong Zhang</i>	

## **ICDM' 06: REGULAR PAPERS**

<b>An Information Theoretic Approach to Detection of Minority Subsets in Database</b> .....	6
<i>Shin Ando, Einoshin Suzuki</i>	
<b>Bayesian State Space Modeling Approach for Measuring the Effectiveness of Marketing Activities and Baseline Sales from POS Data</b> .....	16
<i>Tomohiro Ando</i>	
<b>Learning to Use a Learned Model: A Two-Stage Approach to Classification</b> .....	28
<i>Maria-Luiza Antonie, Osmar Zaiane, Robert Holte</i>	
<b>Hierarchical Classification by Expected Utility Maximization</b> .....	38
<i>Korinna Bade, Eyke Hüllermeier, Andreas Nürnberger</i>	
<b>COALA: A Novel Approach for the Extraction of an Alternate Clustering of High Quality and High Dissimilarity</b> .....	48
<i>Eric Bae, James Bailey</i>	
<b>Cluster Ranking with an Application to Mining Mailbox Networks</b> .....	58
<i>Ziv Bar-Yossef, Ido Guy, Ronny Lempel, Yoëlle Maarek, Vladimir Soroka</i>	
<b>Large Scale Detection of Irregularities in Accounting Data</b> .....	70
<i>Stephen Bay, Krishna Kumaraswamy, Markus Anderle, Rohit Kumar, David Steier</i>	
<b>Adaptive Blocking: Learning to Scale Up Record Linkage</b> .....	82
<i>Mikhail Bilenko, Beena Kamath, Raymond Mooney</i>	
<b>Adaptive Parallel Graph Mining for CMP Architectures</b> .....	92
<i>Gregory Buehrer, Srinivasan Parthasarathy, Yen-Kuang Chen</i>	
<b>Meta Clustering</b> .....	102
<i>Rich Caruana, Mohamed Elhawary, Nam Nguyen, Casey Smith</i>	
<b>Mixed-Drove Spatio-Temporal Co-occurrence Pattern Mining: A Summary of Results</b> .....	114
<i>Mete Celik, Shashi Shekhar, James Rogers, James Shine, Jin Soung Yoo</i>	
<b>An Interactive Semantic Video Mining and Retrieval Platform – Application in Transportation Surveillance Video for Incident Detection</b> .....	124
<i>Xin Chen, Chengcui Zhang</i>	
<b>Tolerance Closed Frequent Itemsets</b> .....	134
<i>James Cheng, Yiping Ke, Wilfred Ng</i>	
<b>Active Learning to Maximize Area Under the ROC Curve</b> .....	144
<i>Matt Culver, Deng Kun, Stephen Scott</i>	

<b>Rapid Identification of Column Heterogeneity .....</b>	<b>154</b>
<i>Bing Tian Dai, Nick Koudas, Beng Chin Ooi, Divesh Srivastava, Suresh Venkatasubramanian</i>	
<b>Data Mining Approaches to Criminal Career Analysis .....</b>	<b>166</b>
<i>Jeroen de Bruin, Tim Cocx, Walter Kusters, Jeroen Laros, Joost Kok</i>	
<b>Biclustering Protein Complex Interactions with a Biclique Finding Algorithm.....</b>	<b>173</b>
<i>Chris Ding, Ya Zhang, Tao Li, Stephen R. Holbrook</i>	
<b>STAGGER: Periodicity Mining of Data Streams Using Expanding Sliding Windows.....</b>	<b>183</b>
<i>Mohamed Elfeky, Walid Aref, Ahmed Elmagarmid</i>	
<b>Turning Clusters into Patterns: Rectangle-based Discriminative Data Description.....</b>	<b>195</b>
<i>Byron Gao, Martin Ester</i>	
<b>Converting Output Scores from Outlier Detection Algorithms into Probability Estimates.....</b>	<b>207</b>
<i>Jing Gao, Pang-Ning Tan</i>	
<b>Personalization in Context: Does Context Matter When Building Personalized Customer Models? .....</b>	<b>217</b>
<i>M. Gorgoglione, C. Palmisano, A. Tuzhilin</i>	
<b>Bregman Bubble Clustering: A Robust, Scalable Framework for Locating Multiple, Dense Regions in Data.....</b>	<b>227</b>
<i>Gunjan Gupta, Joydeep Ghosh</i>	
<b>Optimal Segmentation Using Tree Models .....</b>	<b>239</b>
<i>Robert Gwadera, Aristides Gionis, Heikki Mannila</i>	
<b>Mining for Tree-Query Associations in a Graph.....</b>	<b>249</b>
<i>Eveline Hoekx, Jan Van den Bussche</i>	
<b>Global and Componentwise Extrapolation for Accelerating Data Mining from Large Incomplete Data Sets with the EM Algorithm.....</b>	<b>260</b>
<i>Chun-Nan Hsu, Han-Shen Huang, Bo-Hou Yang</i>	
<b>Keyphrase Extraction Using Semantic Networks Structure Analysis.....</b>	<b>270</b>
<i>Chong Huang, Yonghong Tian, Zhi Zhou, Charles X. Ling, Tiejun Huang</i>	
<b>Subjectivity Categorization of Weblog with Part-of-Speech Based Smoothing.....</b>	<b>280</b>
<i>Shen Huang, Jian-Tao Sun, Xuanhui Wang, Hua-Jun Zeng, Zheng Chen</i>	
<b>Applying Data Mining to Pseudo-Relevance Feedback for High Performance Text Retrieval.....</b>	<b>290</b>
<i>Xiangji Huang, Yan Rui Huang, Miao Wen, Aijun An, Yang Liu, Josiah Poon</i>	
<b>Improving Personalization Solutions through Optimal Segmentation of Customer Bases .....</b>	<b>302</b>
<i>Tianyi Jiang, Alexander Tuzhilin</i>	
<b>Secure Distributed k-Anonymous Pattern Mining.....</b>	<b>314</b>
<i>Wei Jiang, Maurizio Atzori</i>	
<b>Dimension Reduction for Supervised Ordering .....</b>	<b>325</b>
<i>Toshihiro Kamishima, Shotaro Akaho</i>	
<b>A Parameterized Probabilistic Model of Network Evolution for Supervised Link Prediction.....</b>	<b>335</b>
<i>Hisashi Kashima, Naoki Abe</i>	
<b>Incremental Mining of Frequent Query Patterns from XML Queries for Caching .....</b>	<b>345</b>
<i>Guoliang Li, Jianhua Feng, Jianyong Wang, Yong Zhang, Lizhu Zhou</i>	
<b>The Relationships Among Various Nonnegative Matrix Factorization Methods for Clustering.....</b>	<b>357</b>
<i>Tao Li, Chris Ding</i>	

<b>Integrating Features from Different Sources for Music Information Retrieval</b> .....	367
<i>Tao Li, Mitsunori Ogihara, Shenghuo Zhu</i>	
<b>How Bayesians Debug</b> .....	377
<i>Chao Liu, Zeng Lian, Jiawei Han</i>	
<b>On the Use of Structure and Sequence-Based Features for Protein Classification and Retrieval</b> .....	389
<i>Keith Marsolo, Srinivasan Parthasarathy</i>	
<b>Accelerating Newton Optimization for Log-Linear Models through Feature Redundancy</b> .....	399
<i>Arpit Mathur, Soumen Chakrabarti</i>	
<b>P3C: A Robust Projected Clustering Algorithm</b> .....	409
<i>Gabriela Moise, Jörg Sander, Martin Ester</i>	
<b>Frequent Closed Itemset Mining Using Prefix Graphs with an Efficient Flow-based Pruning Strategy</b> .....	421
<i>H. Moonesinghe, Samah Fodeh, Pang-Ning Tan</i>	
<b>Efficient Clustering of Uncertain Data</b> .....	431
<i>Wang Ngai, Ben Kao, Chun Chui, Reynold Cheng, Michael Chau, Kevin Yip</i>	
<b>A Data Mining Approach for Capacity Building of Stakeholders in Integrated Flood Management</b> .....	441
<i>Peter Owotoki, Nataša Manojlovic, Friedrich Mayer-Lindenberg, Erik Pasche</i>	
<b>Local Correlation Tracking in Time Series</b> .....	451
<i>Spiros Papadimitriou, Jimeng Sun, Philip Yu</i>	
<b>Who Thinks Who Knows Who? Socio-cognitive Analysis of Email Networks</b> .....	461
<i>Nishith Pathak, Sandeep Mane, Jaideep Srivastava</i>	
<b>An Efficient Reference-based Approach to Outlier Detection in Large Datasets</b> .....	473
<i>Yaling Pei, Osmar Zaiane, Yong Gao</i>	
<b>Using an Ensemble of One-Class SVM Classifiers to Harden Payload-based Anomaly Detection Systems</b> .....	483
<i>Roberto Perdisci, Guofei Gu, Wenke Lee</i>	
<b>Relational Ensemble Classification</b> .....	494
<i>Christine Preisach, Lars Schmidt-Thieme</i>	
<b>Discovering Partial Orders in Binary Data</b> .....	505
<i>Deepak Rajan, Philip Yu</i>	
<b>Stability Region Based Expectation Maximization for Model-based Clustering</b> .....	517
<i>Chandan Reddy, Hsiao-Dong Chiang, Bala Rajaratnam</i>	
<b>Co-clustering Documents and Words Using Bipartite Isoperimetric Graph Partitioning</b> .....	527
<i>Manjeet Rege, Ming Dong, Farshad Fotouhi</i>	
<b>Latent Dirichlet Co-Clustering</b> .....	537
<i>M. Shafiei, Evangelos Milios</i>	
<b>Latent Friend Mining from Blog Data</b> .....	547
<i>Dou Shen, Jian-Tao Sun, Qiang Yang, Zheng Chen</i>	
<b>The PDD Framework for Detecting Categories of Peculiar Data</b> .....	557
<i>Mahesh Shrestha, Howard Hamilton, Yiyu Yao, Ken Konkel, Liqiang Geng</i>	
<b>Entity Resolution with Markov Logic</b> .....	567
<i>Parag Singla, Pedro Domingos</i>	

<b>Boosting Kernel Models for Regression</b> .....	578
<i>Ping Sun, Xin Yao</i>	
<b>Boosting for Learning Multiple Classes with Imbalanced Class Distribution</b> .....	587
<i>Yanmin Sun, Mohamed Kamel, Yang Wang</i>	
<b>What is the Dimension of your Binary Data?</b> .....	598
<i>Nikolaj Tatti, Taneli Mielikäinen, Aristides Gionis, Heikki Mannila</i>	
<b>Fast Random Walk with Restart and its Applications</b> .....	608
<i>Hanghang Tong, Christos Faloutsos, Jia-Yu Pan</i>	
<b>Anytime Classification Using the Nearest Neighbor Algorithm with Applications to Stream Mining</b> .....	618
<i>Ken Ueno, Xiaopeng Xi, Eamonn Keogh, Dah-Jye Lee</i>	
<b>Dirichlet Aspect Weighting: A Generalized EM Algorithm for Integrating External Data Fields with Semantically Structured Queries by Using Gradient Projection Method</b> .....	628
<i>Atulya Velivelli, Thomas Huang</i>	
<b>Lazy Associative Classification</b> .....	640
<i>Adriano Veloso, Wagner Meira Jr., Mohammed Zakib</i>	
<b>Geometrically Inspired Itemset Mining</b> .....	650
<i>Florian Verhein, Sanjay Chawla</i>	
<b>Finding “Who is Talking to Whom” in VoIP Networks via Progressive Stream Clustering</b> .....	662
<i>Olivier Verscheure, Michail Vlachos, Aris Anagnostopoulos, Pascal Frossard, Eric Bouillet, Philip Yu</i>	
<b>Comparison of Descriptor Spaces for Chemical Compound Retrieval and Classification</b> .....	673
<i>Nikil Wale, George Karypis</i>	
<b>Regularized Least Absolute Deviations Regression and an Efficient Algorithm for Parameter Tuning</b> .....	685
<i>Li Wang, Michael Gordon, Ji Zhu</i>	
<b>LOCI: Load Shedding through Class-Preserving Data Acquisition</b> .....	696
<i>Peng Wang, Haixun Wang, Wei Wang, Baile Shi, Philip Yu</i>	
<b>SAXually Explicit Images: Finding Unusual Shapes</b> .....	706
<i>Li Wei, Eamonn Keogh, Xiaopeng Xi</i>	
<b>A Novel Scalable Algorithm for Supervised Subspace Learning</b> .....	716
<i>Jun Yan, Ning Liu, Benyu Zhang, Qiang Yang, Shuicheng Yan, Zheng Chen</i>	
<b>A Novel Method for Detecting Outlying Subspaces in High-dimensional Databases Using Genetic Algorithm</b> .....	726
<i>Ji Zhang, Qigang Gao, Hai Wang</i>	
<b>Discovering Unrevealed Properties of Probability Estimation Trees: On Algorithm Selection and Performance Explanation</b> .....	736
<i>Kun Zhang, Wei Fan, Bill Buckles, Xiaojing Yuan, Zujia Xu</i>	
<b>Forecasting Skewed Biased Stochastic Ozone Days: Analyses and Solutions</b> .....	748
<i>Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, Xiangshang Li</i>	
<b>Identifying Follow-Correlation Itemset-Pairs</b> .....	760
<i>Shichao Zhang, Jilian Zhang, Xiaofeng Zhu, Zifang Huang</i>	
<b>On the Lower Bound of Local Optimums in K-Means Algorithm</b> .....	770
<i>Zhenjie Zhang, Bing Tian Dai, Anthony Tung</i>	

## **ICDM' 06: SHORT PAPERS**

<b>Fast On-line Kernel Learning for Trees</b> .....	780
<i>Fabio Aioli, Giovanni Da San Martino, Alessandro Sperduti, Alessandro Moschitti</i>	
<b>bitSPADE: A Lattice-based Sequential Pattern Mining Algorithm Using Bitmap Representation</b> .....	785
<i>Sujeewan Aseervatham, Aomar Osmani, Emmanuel Viennet</i>	
<b>Decision Trees for Functional Variables</b> .....	791
<i>Suhrid Balakrishnan, David Madigan</i>	
<b>Mining Latent Associations of Objects Using a Typed Mixture Model — A Case Study on Expert/Expertise Mining</b> .....	796
<i>Shenghua Bao, Yunbo Cao, Bing Liu, Yong Yu, Hang Li</i>	
<b>Semantic Kernels for Text Classification Based on Topological Measures of Feature Similarity</b> .....	801
<i>Stephan Bloehdorn, Roberto Basili, Marco Cammisa, Alessandro Moschitti</i>	
<b>Mining Maximal Generalized Frequent Geographic Patterns with Knowledge Constraints</b> .....	806
<i>Vania Bogorny, João Valiati, Sandro Camargo, Paulo Engel, Bart Kuijpers, Luis Alvares</i>	
<b>Pattern Mining in Frequent Dynamic Subgraphs</b> .....	811
<i>Karsten Borgwardt, Hans-Peter Kriegel, Peter Wackersreuther</i>	
<b>Discovery of Collocation Episodes in Spatiotemporal Data</b> .....	816
<i>Huiping Cao, Nikos Mamoulis, David Cheung</i>	
<b>Getting the Most Out of Ensemble Selection</b> .....	821
<i>Rich Caruana, Art Munson, Alexandru Niculescu-Mizil</i>	
<b>Diverse Topic Phrase Extraction through Latent Semantic Analysis</b> .....	827
<i>Jilin Chen, Jun Yan, Benyu Zhang, Qiang Yang, Zheng Chen</i>	
<b>AC-Close: Efficiently Mining Approximate Closed Itemsets by Core Pattern Recovery</b> .....	832
<i>Hong Cheng, Philip Yu, Jiawei Han</i>	
<b>Belief Propagation in Large, Highly Connected Graphs for 3D Part-Based Object Recognition</b> .....	838
<i>Frank DiMaio, Jude Shavlik</i>	
<b>A Framework for Regional Association Rule Mining in Spatial Datasets</b> .....	844
<i>Wei Ding, Christoph Eick, Jing Wang, Xiaojing Yuan</i>	
<b>Mining Generalized Graph Patterns Based on User Examples</b> .....	850
<i>Pavel Dmitriev, Carl Lagoze</i>	
<b>An Experimental Investigation of Graph Kernels on a Collaborative Recommendation Task</b> .....	856
<i>Francois Fouss, Luh Yen, Alain Pirotte, Marco Saerens</i>	
<b>A Balanced Ensemble Approach to Weighting Classifiers for Text Classification</b> .....	862
<i>Gabriel Fung, Jeffrey Yu, Haixun Wang, David Cheung, Huan Liu</i>	
<b>Detection of Interdomain Routing Anomalies Based on Higher-Order Path Analysis</b> .....	867
<i>Murat Can Ganiz, Sudhan Kanitkar, Mooi Choo Chuah, William Pottenger</i>	
<b>Star-Structured High-Order Heterogeneous Data Co-clustering Based on Consistent Information Theory</b> .....	873
<i>Bin Gao, Tie-Yan Liu, Wei-Ying Ma</i>	

<b>GraphRank: Statistical Modeling and Mining of Significant Subgraphs in the Feature Space</b> .....	878
<i>Huahai He, Ambuj Singh</i>	
<b>A Feature Selection and Evaluation Scheme for Computer Virus Detection</b> .....	884
<i>Olivier Henchiri, Nathalie Japkowicz</i>	
<b>Cluster Analysis of Time-series Medical Data Based on the Trajectory Representation and Multiscale Comparison Techniques</b> .....	889
<i>Shoji Hirano, Shusaku Tsumoto</i>	
<b>Constructing Ensembles for Better Ranking</b> .....	895
<i>Jin Huang, Charles X. Ling</i>	
<b>TRIAS — An Algorithm for Mining Iceberg Tri-Lattices</b> .....	900
<i>Robert Jäschke, Andreas Hotho, Christoph Schmitz, Bernhard Ganter, Gerd Stumme</i>	
<b>Intelligent Icons: Integrating Lite-Weight Data Mining and Visualization into GUI Operating Systems</b> .....	905
<i>Eamonn Keogh, Li Wei, Xiaopeng Xi, Stefano Lonardi, Jin Shieh, Scott Sirowy</i>	
<b>COSMIC: Conceptually Specified Multi-Instance Clusters</b> .....	910
<i>Hans-Peter Kriegel, Alexey Pryakhin, Matthias Schubert, Arthur Zimek</i>	
<b>Direct Marketing When There Are Voluntary Buyers</b> .....	915
<i>Yi-Ting Lai, Ke Wang, Daymond Ling, Hua Shi, Jason Zhang</i>	
<b>DSTree: A Tree Structure for the Mining of Frequent Sets from Data Streams</b> .....	921
<i>Carson Leung, Quamrul Khan</i>	
<b>Searching for Pattern Rules</b> .....	926
<i>Guichong Li, Howard Hamilton</i>	
<b>Adding Semantics to Email Clustering</b> .....	931
<i>Hua Li, Dou Shen, Benyu Zhang, Zheng Chen, Qiang Yang</i>	
<b>Gradual Cube: Customize Profile on Mobile OLAP</b> .....	936
<i>Jun Li, Haofeng Zhou, Wei Wang</i>	
<b>CoMiner: An Effective Algorithm for Mining Competitors from the Web</b> .....	941
<i>Rui Li, Shenghua Bao, Jin Wang, Yong Yu, Yunbo Cao</i>	
<b>Multi-Tier Granule Mining for Representations of Multidimensional Association Rules</b> .....	946
<i>Yuefeng Li, Wanzhong Yang, Yue Xu</i>	
<b>Social Capital in Friendship-Event Networks</b> .....	952
<i>Louis Licamele, Lise Getoor</i>	
<b>Exploratory Under-Sampling for Class-Imbalance Learning</b> .....	958
<i>Xu-Ying Liu, Jianxin Wu, Zhi-Hua Zhou</i>	
<b>The Influence of Class Imbalance on Cost-Sensitive Learning: An Empirical Study</b> .....	963
<i>Xu-Ying Liu, Zhi-Hua Zhou</i>	
<b>Similarity of Temporal Query Logs Based on ARIMA Model</b> .....	968
<i>Ning Liu, Shuzhen Nong, Jun Yan, Benyu Zhang, Zheng Chen, Ying Li</i>	
<b>Probabilistic Segmentation and Analysis of Horizontal Cells</b> .....	973
<i>Vebjorn Ljosa, Ambuj Singh</i>	
<b>Mining Correlation between Motifs and Gene Expression</b> .....	979
<i>Yi Lu, Shiyong Lu, Adrian Platts, Stephen Krawetz</i>	

<b>High Quality, Efficient Hierarchical Document Clustering Using Closed Interesting Itemsets</b> .....	984
<i>Hassan Malik, John Kender</i>	
<b>On Trajectory Representation for Scientific Features</b> .....	990
<i>Sameep Mehta, Srinivasan Parthasarathy, Raghu Machiraju</i>	
<b>NewsCATS: A News Categorization and Trading System</b> .....	995
<i>Marc-André Mittermayer, Gerhard Knolmayer</i>	
<b>Improving Grouped-Entity Resolution Using Quasi-Cliques</b> .....	1001
<i>Byung-Won On, Ergin Elmacioglu, Dongwon Lee, Jaewoo Kang, Jian Pei</i>	
<b>Fast Relevance Discovery in Time Series</b> .....	1009
<i>Chang-shing Perng, Haixun Wang, Sheng Ma</i>	
<b>Probabilistic Enhanced Mapping with the Generative Tabular Model</b> .....	1014
<i>Rodolphe Priam, Mohamed Nadif</i>	
<b>Object Identification with Constraints</b> .....	1019
<i>Steffen Rendle, Lars Schmidt-Thieme</i>	
<b>High-Performance Unsupervised Relation Extraction from Large Corpora</b> .....	1025
<i>Binjamin Rozenfeld, Ronen Feldman</i>	
<b>Cluster Based Core Vector Machine</b> .....	1031
<i>Asharaf S., M. Murty, S. Shevade</i>	
<b>Enhancing Text Clustering Using Concept-based Mining Model</b> .....	1036
<i>Shady Shehata, Fakhri Karray, Mohamed Kamel</i>	
<b>Detecting Link Spam using Temporal Information</b> .....	1042
<i>Guoyang Shen, Bin Gao, Tie-Yan Liu, Guang Feng, Shiji Song, Hang Li</i>	
<b>Minimum Enclosing Spheres Formulations for Support Vector Ordinal Regression</b> .....	1047
<i>S. Shevade, Wei Chu</i>	
<b>Mining Maximal Quasi-Bicliques to Co-Cluster Stocks and Financial Ratios for Value Investment</b> .....	1052
<i>Kelvin Sim, Jinyan Li, Vivekanand Gopalkrishnan, Guimei Liu</i>	
<b>Boosting the Feature Space: Text Classification for Unstructured Data on the Web</b> .....	1057
<i>Yang Song, Ding Zhou, Jian Huang, Isaac Councill, Hongyuan Zha, C. Giles</i>	
<b>Plagiarism Detection in arXiv</b> .....	1063
<i>Daria Sorokina, Johannes Gehrke, Simeon Warner, Paul Ginsparg</i>	
<b>Window-based Tensor Analysis on High-dimensional and Multi-aspect Streams</b> .....	1069
<i>Jimeng Sun, Spiros Papadimitriou, Philip Yu</i>	
<b>Automatic Single-Organ Segmentation in Computed Tomography Images</b> .....	1074
<i>Ruchaneewan Susomboon, Daniela Raicu, Jacob Furst, David Chamin</i>	
<b>Improving Nearest Neighbor Classifier Using Tabu Search and Ensemble Distance Metrics</b> .....	1079
<i>Muhammad Tahir, James Smith</i>	
<b>Comparisons of K-Anonymization and Randomization Schemes Under Linking Attacks</b> .....	1084
<i>Zhouxuan Teng, Wenliang Du</i>	
<b>MARGIN: Maximal Frequent Subgraph Mining</b> .....	1090
<i>Lini Thomas, Satyanarayana Valluri, Kamalakar Karlapalem</i>	
<b>Resource Management for Networked Classifiers in Distributed Stream Mining Systems</b> .....	1095
<i>Deepak Turaga, Olivier Verscheure, Upendra Chaudhari, Lisa Amini</i>	



<b>A Simple Yet Effective Data Clustering Algorithm</b> .....	1101
<i>Soujanya Vadapalli, Satyanarayana Valluri, Kamalakar Karlapalem</i>	
<b>Entropy-based Concept Shift Detection</b> .....	1106
<i>Peter Vorburget, Abraham Bernstein</i>	
<b>Recommendation on Item Graphs</b> .....	1112
<i>Fei Wang, Sheng Ma, Liuzhong Yang, Tao Li</i>	
<b>Solution Path for Semi-Supervised Classification with Manifold Regularization</b> .....	1117
<i>Gang Wang, Tao Chen, Dit-Yan Yeung, Frederick Lochofsky</i>	
<b>Semi-Supervised Kernel Regression</b> .....	1123
<i>Meng Wang, Xian-Sheng Hua, Yan Song, Li-Rong Dai, Hong-Jiang Zhang</i>	
<b>Mining Complex Time-Series Data by Learning Markovian Models</b> .....	1129
<i>Yi Wang, Lizhu Zhou, Jianhua Feng, Jianyong Wang, Zhi-Qiang Liu</i>	
<b>Temporal Data Mining in Dynamic Feature Spaces</b> .....	1134
<i>Brent Wenerstrom, Christophe Giraud-Carrier</i>	
<b>Discover Bayesian Networks from Incomplete Data Using a Hybrid Evolutionary Algorithm</b> .....	1139
<i>Man Leung Wong, Yuan Yuan Guo</i>	
<b>Distances and (Indefinite) Kernels for Sets of Objects</b> .....	1144
<i>Adam Woznica, Alexandros Kalousis, Melanie Hilario</i>	
<b>Deploying Approaches for Pattern Refinement in Text Mining</b> .....	1150
<i>Sheng-Tang Wu, Yuefeng Li, Yue Xu</i>	
<b>TOP-COP: Mining TOP-K Strongly Correlated Pairs in Large Databases</b> .....	1155
<i>Hui Xiong, Mark Brodie, Sheng Ma</i>	
<b>Manifold Clustering of Shapes</b> .....	1160
<i>Dragomir Yankov, Eamonn Keogh</i>	
<b>Linear and Non-Linear Dimensional Reduction via Class Representatives for Text Classification</b> .....	1165
<i>Dimitrios Zeimpekis, Efstratios Gallopoulos</i>	
<b>Adaptive Kernel Principal Component Analysis with Unsupervised Learning of Kernels</b> .....	1171
<i>Daoqiang Zhang, Zhi-Hua Zhou, Songcan Chen</i>	
<b>Rule-Based Platform for Web User Profiling</b> .....	1176
<i>Jianping Zhang, Manu Shukla</i>	
<b>Opening the Black Box of Feature Extraction: Incorporating Visualization into High-Dimensional Data Mining Processes</b> .....	1181
<i>Jianting Zhang, Le Gruenwald</i>	
<b>Semantic Smoothing for Model-based Document Clustering</b> .....	1186
<i>Xiaodan Zhang, Xiaohua Zhou, Xiaohua Hu</i>	
<b>Corrective Classification: Classifier Ensembling with Corrective and Diverse Base Learners</b> .....	1192
<i>Yan Zhang, Xingquan Zhu, Xindong Wu</i>	
<b>Speedup Clustering with Hierarchical Ranking</b> .....	1198
<i>Jianjun Zhou, Joerg Sander</i>	
<b>Query-Sensitive Similarity Measure for Content-Based Image Retrieval</b> .....	1204
<i>Zhi-Hua Zhou, Hong-Bin Dai</i>	
<b>Author Index</b>	