

2009 9th IEEE International Conference on Data Mining (ICDM 2009)

**Miami Beach, Florida, USA
6 – 9 December 2009**

Pages 1-542



**IEEE Catalog Number: CFP09278-PRT
ISBN: 978-1-4244-5242-2**

2009 Ninth IEEE International Conference on Data Mining

ICDM 2009

Table of Contents

Message from the General Co-Chairs
Message from the Program Committee Co-Chairs
Organizing Committee
Program Committee
Steering Committee

Regular Papers

Explore/Exploit Schemes for Web Content Optimization	1
<i>Deepak Agarwal, Bee-Chung Chen, and Pradheep Elango</i>	
Connecting Sparsely Distributed Similar Bloggers	11
<i>Nitin Agarwal, Huan Liu, Shankara Subramanya, John J. Salerno, and Philip S. Yu</i>	
Rule Ensembles for Multi-target Regression	21
<i>Timo Aho, Bernard Ženko, and Sašo Džeroski</i>	
A Local Scalable Distributed Expectation Maximization Algorithm for Large Peer-to-Peer Networks	31
<i>Kanishka Bhaduri and Ashok N. Srivastava</i>	
Cross-Guided Clustering: Transfer of Relevant Supervision across Domains for Improved Clustering	41
<i>Indrajit Bhattacharya, Shantanu Godbole, Sachindra Joshi, and Ashish Verma</i>	
Audio Classification of Bird Species: A Statistical Manifold Approach	51
<i>Forrest Briggs, Raviv Raich, and Xiaoli Z. Fern</i>	
Finding Associations and Computing Similarity via Biased Pair Sampling	61
<i>Andrea Campagna and Rasmus Pagh</i>	
Beyond Banditron: A Conservative and Efficient Reduction for Online Multiclass Prediction with Bandit Setting Model	71
<i>Guangyun Chen, Gang Chen, Jianwen Zhang, Shuo Chen, and Changshui Zhang</i>	

Probabilistic Similarity Query on Dimension Incomplete Data	81
<i>Wei Cheng, Xiaoming Jin, and Jian-Tao Sun</i>	
flowNet: Flow-Based Approach for Efficient Analysis of Complex Biological Networks	91
<i>Young-Rae Cho, Lei Shi, and Aidong Zhang</i>	
v-Anomica: A Fast Support Vector Based Novelty Detection Technique	101
<i>Santanu Das, Kanishka Bhaduri, Nikunj C. Oza, and Ashok N. Srivastava</i>	
Temporal Neighborhood Discovery Using Markov Models	110
<i>Sandipan Dey, Vandana P. Janeja, and Aryya Gangopadhyay</i>	
Active Learning with Generalized Queries	120
<i>Jun Du and Charles X. Ling</i>	
Conditional Models for Non-smooth Ranking Loss Functions	129
<i>Avinava Dubey, Jinesh Machchhar, Chiranjib Bhattacharyya, and Soumen Chakrabarti</i>	
Unsupervised Class Separation of Multivariate Data through Cumulative Variance-Based Ranking	139
<i>Andrew Foss, Osmar R. Zaiane, and Sandra Zilles</i>	
Execution Anomaly Detection in Distributed Systems through Unstructured Log Analysis	149
<i>Qiang Fu, Jian-Guang Lou, Yi Wang, and Jiang Li</i>	
Learning the Shared Subspace for Multi-task Clustering and Transductive Transfer Classification	159
<i>Quanquan Gu and Jie Zhou</i>	
Accurate Estimation of the Degree Distribution of Private Networks	169
<i>Michael Hay, Chao Li, Jerome Miklau, and David Jensen</i>	
A Linear-Time Graph Kernel	179
<i>Shohei Hido and Hisashi Kashima</i>	
GSML: A Unified Framework for Sparse Metric Learning	189
<i>Kaizhu Huang, Yiming Ying, and Colin Campbell</i>	
GRAPE: A Graph-Based Framework for Disambiguating People Appearances in Web Search	199
<i>Lili Jiang, Jianyong Wang, Ning An, Shengyuan Wang, Jian Zhan, and Lian Li</i>	
A Tree-Based Framework for Difference Summarization	209
<i>Ruoming Jin, Yuri Breitbart, and Rong Li</i>	
TrBagg: A Simple Transfer Learning Method and its Application to Personalization in Collaborative Tagging	219
<i>Toshihiro Kamishima, Masahiro Hamasaki, and Shotaro Akaho</i>	
PEGASUS: A Peta-Scale Graph Mining System—Implementation and Observations	229
<i>U. Kang, Charalampos E. Tsourakakis, and Christos Faloutsos</i>	
Efficient Discovery of Frequent Correlated Subgraph Pairs	239
<i>Yiping Ke, James Cheng, and Jeffrey Xu Yu</i>	

Self-adaptive Anytime Stream Clustering	249
<i>Philipp Kranen, Ira Assent, Corinna Baldauf, and Thomas Seidl</i>	
Improving SVM Classification on Imbalanced Data Sets in Distance Spaces	259
<i>Suzan Köknar-Tezel and Longin Jan Latecki</i>	
CoCoST: A Computational Cost Efficient Classifier	268
<i>Liyun Li, Umut Topkara, Baris Coskun, and Nasir Memon</i>	
Semi-naive Exploitation of One-Dependence Estimators	278
<i>Nan Li, Yang Yu, and Zhi-Hua Zhou</i>	
A Framework for Computing the Privacy Scores of Users in Online Social Networks	288
<i>Kun Liu and Evimaria Terzi</i>	
Least Square Incremental Linear Discriminant Analysis	298
<i>Li-Ping Liu, Yuan Jiang, and Zhi-Hua Zhou</i>	
Unified Solution to Nonnegative Data Factorization Problems	307
<i>Xiaobai Liu, Shuicheng Yan, Jun Yan, and Hai Jin</i>	
Extended Boolean Matrix Decomposition	317
<i>Haibing Lu, Jaideep Vaidya, Vijayalakshmi Atluri, and Yuan Hong</i>	
Active Learning with Adaptive Heterogeneous Ensembles	327
<i>Zhenyu Lu, Xindong Wu, and Josh Bongard</i>	
Non-negative Laplacian Embedding	337
<i>Dijun Luo, Chris Ding, Heng Huang, and Tao Li</i>	
Scalable Algorithms for Distribution Search	347
<i>Yasuko Matsubara, Yasushi Sakurai, and Masatoshi Yoshikawa</i>	
A Deep Non-linear Feature Mapping for Large-Margin kNN Classification	357
<i>Renqiang Min, David A. Stanley, Zineng Yuan, Anthony Bonner, and Zhaolei Zhang</i>	
Finding Time Series Motifs in Disk-Resident Data	367
<i>Abdullah Mueen, Eamonn Keogh, and Nima Bigdely-Shamlo</i>	
Relevant Subspace Clustering: Mining the Most Interesting Non-redundant Concepts in High Dimensional Data	377
<i>Emmanuel Müller, Ira Assent, Stephan Günnemann, Ralph Krieger, and Thomas Seidl</i>	
Stacked Gaussian Process Learning	387
<i>Marion Neumann, Kristian Kersting, Zhao Xu, and Daniel Schulz</i>	
Evaluating Statistical Tests for Within-Network Classifiers of Relational Data	397
<i>Jennifer Neville, Brian Gallagher, and Tina Eliassi-Rad</i>	
Discovering Excitatory Networks from Discrete Event Streams with Applications to Neuronal Spike Train Analysis	407
<i>Debprakash Patnaik, Srivatsan Laxman, and Naren Ramakrishnan</i>	
Clustering Trajectories of Moving Objects in an Uncertain World	417
<i>Nikos Pelekis, Ioannis Kopanakis, Evangelos Kotsifakos, Elias Frentzos, and Yannis Theodoridis</i>	

Semi-Supervised Sequence Labeling with Self-Learned Features	428
<i>YanJun Qi, Pavel Kuksa, Ronan Collobert, Kunihiko Sadamasa, Koray Kavukcuoglu, and Jason Weston</i>	
Semi-Markov kMeans Clustering and Activity Recognition from Body-Worn Sensors	438
<i>Matthew Robards and Peter Sunehag</i>	
A Sparsification Approach for Temporal Graphical Model Decomposition	447
<i>Ning Ruan, Ruoming Jin, Victor E. Lee, and Kun Huang</i>	
Resolving Identity Uncertainty with Learned Random Walks	457
<i>Ted Sandler, Lyle H. Ungar, and Koby Crammer</i>	
Discriminative Mixed-Membership Models	466
<i>Hanhui Shan, Arindam Banerjee, and Nikunj C. Oza</i>	
Argumentation Based Constraint Acquisition	476
<i>Kostyantyn Shchekotykhin and Gerhard Friedrich</i>	
Extending Semi-supervised Learning Methods for Inductive Transfer Learning	483
<i>Yuan Shi, Zhenzhong Lan, Wei Liu, and Wei Bi</i>	
iTopicModel: Information Network-Integrated Topic Modeling	493
<i>Yizhou Sun, Jiawei Han, Jing Gao, and Yintao Yu</i>	
Uncovering Groups via Heterogeneous Interaction Analysis	503
<i>Lei Tang, Xufei Wang, and Huan Liu</i>	
Significance of Episodes Based on Minimal Windows	513
<i>Nikolaj Tatti</i>	
Convex Non-negative Matrix Factorization in the Wild	523
<i>Christian Thureau, Kristian Kersting, and Christian Bauckhage</i>	
On K-Means Cluster Preservation Using Quantization Schemes	533
<i>Deepak S. Turaga, Michail Vlachos, and Olivier Verscheure</i>	
Scalable Classification in Large Scale Spatiotemporal Domains Applied to Voltage-Sensitive Dye Imaging	543
<i>Igor Vainer, Sarit Kraus, Gal Kaminka, and Hamutal Slovin</i>	
Extracting Output Metadata from Scientific Deep Web Data Sources	552
<i>Fan Wang and Gagan Agrawal</i>	
Semi-supervised Multi-task Learning with Task Regularizations	562
<i>Fei Wang, Xin Wang, and Tao Li</i>	
Fast Online Training of Ramp Loss Support Vector Machines	569
<i>Zhuang Wang and Slobodan Vucetic</i>	
Mining Peculiarity Groups in Day-by-Day Behavioral Datasets	578
<i>Yun Xiong and Yangyong Zhu</i>	
Online System Problem Detection by Mining Patterns of Console Logs	588
<i>Wei Xu, Ling Huang, Armando Fox, David Patterson, and Michael Jordan</i>	

Synthesizing Novel Dimension Reduction Algorithms in Matrix Trace Oriented Optimization Framework	598
<i>Jun Yan, Ning Liu, Shuicheng Yan, Qiang Yang, and Zheng Chen</i>	
Peculiarity Analysis for Classifications	607
<i>Jian Yang, Ning Zhong, Yiyu Yao, and Jue Wang</i>	
Filtering and Refinement: A Two-Stage Approach for Efficient and Effective Anomaly Detection	617
<i>Xiao Yu, Lu An Tang, and Jiawei Han</i>	
Mining Data Streams with Labeled and Unlabeled Training Examples	627
<i>Peng Zhang, Xingquan Zhu, and Li Guo</i>	
Maximum Margin Clustering with Multivariate Loss Function	637
<i>Bin Zhao, James Kwok, and Changshui Zhang</i>	
Efficient Discovery of Confounders in Large Data Sets	647
<i>Wenjun Zhou and Hui Xiong</i>	
Vague One-Class Learning for Data Streams	657
<i>Xingquan Zhu, Xindong Wu, and Chengqi Zhang</i>	
Inverse Time Dependency in Convex Regularized Learning	667
<i>Zeyuan Allen Zhu, Weizhu Chen, Chenguang Zhu, Gang Wang, Haixun Wang, and Zheng Chen</i>	
P-packSVM: Parallel Primal grAdient desCent Kernel SVM	677
<i>Zeyuan Allen Zhu, Weizhu Chen, Gang Wang, Chenguang Zhu, and Zheng Chen</i>	
Short Papers	
An L-infinity Norm Visual Classifier	687
<i>Anushka Anand, Leland Wilkinson, and Dang Nhon Tuan</i>	
Outlier Detection Using Inductive Logic Programming	693
<i>Fabrizio Angiulli and Fabio Fassetti</i>	
Joint Emotion-Topic Modeling for Social Affective Text Mining	699
<i>Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, and Yong Yu</i>	
Algorithms for Large, Sparse Network Alignment Problems	705
<i>Mohsen Bayati, Margot Gerritsen, David F. Gleich, Amin Saberi, and Ying Wang</i>	
Dirichlet Mixture Allocation for Multiclass Document Collections Modeling	711
<i>Wei Bian and Dacheng Tao</i>	
SLIDER: Mining Correlated Motifs in Protein-Protein Interaction Networks	716
<i>Peter Boyen, Frank Neven, Dries Van Dyck, Aalt-Jan van Dijk, and Roeland C.H.J. van Ham</i>	
Effective Anomaly Detection in Sensor Networks Data Streams	722
<i>Saha Budhaditya, Duc-Son Pham, Mihai Lazarescu, and Svetha Venkatesh</i>	
Hierarchical Bayesian Models for Collaborative Tagging Systems	728
<i>Markus Bundschuh, Shipeng Yu, Volker Tresp, Achim Rettinger, Mathaeus Dejori, and Hans-Peter Kriegel</i>	

Efficient Algorithm for Computing Link-Based Similarity in Real World Networks	734
<i>Yuanzhe Cai, Gao Cong, Xu Jia, Hongyan Liu, Jun He, Jiaheng Lu, and Xiaoyong Du</i>	
Fine-Grain Perturbation for Privacy Preserving Data Publishing	740
<i>Rhonda Chaytor, Ke Wang, and Patricia Brantingham</i>	
Accelerated Gradient Method for Multi-task Sparse Learning Problem	746
<i>Xi Chen, Weike Pan, James T. Kwok, and Jaime G. Carbonell</i>	
CoFKM: A Centralized Method for Multiple-View Clustering	752
<i>Guillaume Cleuziou, Matthieu Exbrayat, Lionel Martin, and Jacques-Henri Sublemontier</i>	
Active Selection of Sensor Sites in Remote Sensing Applications	758
<i>Debasish Das, Zoran Obradovic, and Slobodan Vucetic</i>	
Large Scale Relation Acquisition Using Class Dependent Patterns	764
<i>Stijn De Saeger, Kentaro Torisawa, Jun'ichi Kazama, Kow Kuroda, and Masaki Murata</i>	
Finding Maximal Fully-Correlated Itemsets in Large Databases	770
<i>Lian Duan and William Nick Street</i>	
Bayesian Overlapping Subspace Clustering	776
<i>Qiang Fu and Arindam Banerjee</i>	
Unsupervised Relation Extraction by Massive Clustering	782
<i>Edgar González and Jordi Turmo</i>	
Regression Learning Vector Quantization	788
<i>Mihajlo Grbovic and Slobodan Vucetic</i>	
Projective Clustering Ensembles	794
<i>Francesco Gullo, Carlotta Domeniconi, and Andrea Tagarelli</i>	
Knowledge Discovery from Citation Networks	800
<i>Zhen Guo, Zhongfei Zhang, Shenghuo Zhu, Yun Chi, and Yihong Gong</i>	
An Effective Approach to Inverse Frequent Set Mining	806
<i>Antonella Guzzo, Domenico Saccà, and Edoardo Serra</i>	
Parallel PathFinder Algorithms for Mining Structures from Graphs	812
<i>Samson Hauguel, ChengXiang Zhai, and Jiawei Han</i>	
A Bootstrap Approach to Eigenvalue Correction	818
<i>Anne Hendrikse, Luuk Spreeuwiers, and Raymond Veldhuis</i>	
Modeling Syntactic Structures of Topics with a Nested HMM-LDA	824
<i>Jing Jiang</i>	
Redistricting Using Heuristic-Based Polygonal Clustering	830
<i>Deepti Joshi, Leen-Kiat Soh, and Ashok Samal</i>	
A Walk from 2-Norm SVM to 1-Norm SVM	836
<i>Jussi Kujala, Timo Aho, and Tapio Elomaa</i>	
Semi-supervised Density-Based Clustering	842
<i>Levi Lelis and Jörg Sander</i>	
VIF Regression: A Fast Regression Algorithm for Large Data	848
<i>Dongyu Lin and Dean P. Foster</i>	

Sparse Norm-Regularized Reconstructive Coefficients Learning	854
<i>Bin Liu, Shuo Chen, Mingjie Qian, and Changshui Zhang</i>	
A Contrast Pattern Based Clustering Quality Index for Categorical Data	860
<i>Qingbao Liu and Guozhu Dong</i>	
Multi-document Summarization by Information Distance	866
<i>Chong Long, Minlie Huang, Xiaoyan Zhu, and Ming Li</i>	
On the (In)Security and (Im)Practicality of Outsourcing Precise Association Rule Mining	872
<i>Ian Molloy, Ninghui Li, and Tiancheng Li</i>	
Promoting Total Efficiency in Text Clustering via Iterative and Interactive Metric Learning	878
<i>Michinari Momma, Satoshi Morinaga, and Daisuke Komura</i>	
A New Clustering Algorithm Based on Regions of Influence with Self-Detection of the Best Number of Clusters	884
<i>Fabrice Muhlenbach and Stéphane Lallich</i>	
Automatically Extracting Dialog Models from Conversation Transcripts	890
<i>Sumit Negi, Sachindra Joshi, Anup K. Chalamalla, and L. Venkata Subramaniam</i>	
To Trust or Not to Trust? Predicting Online Trusts Using Trust Antecedent Framework	896
<i>Viet-An Nguyen, Ee-Peng Lim, Jing Jiang, and Aixin Sun</i>	
Analysis of Subsequence Time-Series Clustering Based on Moving Average	902
<i>Miho Ohsaki, Masakazu Nakase, and Shigeru Katagiri</i>	
Permutation Tests for Studying Classifier Performance	908
<i>Markus Ojala and Gemma C. Garriga</i>	
Interaction-Based Clustering of Multivariate Time Series	914
<i>Claudia Plant, Afra M. Wohlschläger, and Andrew Zherdin</i>	
PUB: A Class Description Technique Based on Partial Coverage of Subspace	920
<i>Ardian Kristanto Poernomo and Vivekanand Gopalkrishnan</i>	
Online and Batch Learning of Generalized Cosine Similarities	926
<i>Ali Mustafa Qamar and Eric Gaussier</i>	
Discovering Organizational Structure in Dynamic Social Network	932
<i>Jiangtao Qiu, Zhangxi Lin, Changjie Tang, and Shaojie Qiao</i>	
Kernel Conditional Quantile Estimation via Reduction Revisited	938
<i>Novi Quadrianto, Kristian Kersting, Mark D. Reid, Tibério S. Caetano, and Wray L. Buntine</i>	
Naïve Bayes Classification of Uncertain Data	944
<i>Jiangtao Ren, Sau Dan Lee, Xianlu Chen, Ben Kao, Reynold Cheng, and David Cheung</i>	
Constraint-Based Pattern Mining in Dynamic Graphs	950
<i>Céline Robardet</i>	
Global Slope Change Synopses for Measurement Maps	956
<i>Frank Rosenthal, Ulrike Fischer, Peter B. Volk, and Wolfgang Lehner</i>	

Aspect Guided Text Categorization with Unobserved Labels	962
<i>Dan Roth and Yuancheng Tu</i>	
A Fully Automated Method for Discovering Community Structures in High Dimensional Data	968
<i>Jianhua Ruan</i>	
Hierarchical Probabilistic Segmentation of Discrete Events	974
<i>Guy Shani, Christopher Meek, and Asela Gunawardana</i>	
Topic Modeling for Sequences of Temporal Activities	980
<i>Zhiyong Shen, Ping Luo, Yuhong Xiong, Jun Sun, and Yidong Shen</i>	
Combining Super-Structuring and Abstraction on Sequence Classification	986
<i>Adrian Silvescu, Cornelia Caragea, and Vasant Honavar</i>	
A Global-Model Naive Bayes Approach to the Hierarchical Prediction of Protein Functions	992
<i>Carlos N. Silla Jr. and Alex A. Freitas</i>	
Spatio-temporal Energy Based Gait Recognition	998
<i>Shamsher Singh and K.K. Biswas</i>	
Feature Selection in the Tensor Product Feature Space	1004
<i>Aaron Smalter, Jun Huan, and Gerald Lushington</i>	
Topic Distributions over Links on Web	1010
<i>Jie Tang, Jing Zhang, Jeffrey Xu Yu, Zi Yang, Keke Cai, Rui Ma, Li Zhang, and Zhong Su</i>	
Clustering with Multiple Graphs	1016
<i>Wei Tang, Zhengdong Lu, and Inderjit S. Dhillon</i>	
Two Heads Better Than One: Metric+Active Learning and its Applications for IT Service Classification	1022
<i>Fei Wang, Jimeng Sun, Tao Li, and Nikos Anerousis</i>	
Maximum Margin Clustering on Data Manifolds	1028
<i>Fei Wang, Xin Wang, and Tao Li</i>	
Discovering Contexts and Contextual Outliers Using Random Walks in Graphs	1034
<i>Xiang Wang and Ian Davidson</i>	
Effective Criterion Functions for Efficient Agglomerative Clustering on Very Large Networks	1040
<i>Yang Wang and Mingyuan An</i>	
Binomial Matrix Factorization for Discrete Collaborative Filtering	1046
<i>Jinlong Wu</i>	
Bi-relational Network Analysis Using a Fast Random Walk with Restart	1052
<i>Jing Xia, Doina Caragea, and William H. Hsu</i>	
A New MCA-Based Divisive Hierarchical Algorithm for Clustering Categorical Data	1058
<i>Tengke Xiong, Shengrui Wang, André Mayers, and Ernest Monga</i>	

Non-sparse Multiple Kernel Learning for Fisher Discriminant Analysis	1064
<i>Fei Yan, Josef Kittler, Krystian Mikolajczyk, and Atif Tahir</i>	
Multirelational Topic Models	1070
<i>Jia Zeng, William K. Cheung, Chun-hung Li, and Jiming Liu</i>	
Learning Local Components to Understand Large Bayesian Networks	1076
<i>Yifeng Zeng, Yanping Xiang, Jorge Cordero H., and Yujian Lin</i>	
RING: An Integrated Method for Frequent Representative Subgraph Mining	1082
<i>Shijie Zhang, Jiong Yang, and Shirong Li</i>	
A Cost-Effective LSH Filter for Fast Pairwise Mining	1088
<i>Gang Zhao, Yun Xiong, Longbing Cao, Dan Luo, Xuchun Su, and Yangyong Zhu</i>	
A New Kernel-Based Classification Algorithm	1094
<i>Xiaofei Zhou, Wenhan Jiang, Yingjie Tian, Peng Zhang, Guangli Nie, and Yong Shi</i>	

Author Index