

11th SIAM International Conference on Data Mining 2011

**Mesa, Arizona, USA
28-30 April 2011**

Volume 1 of 2

ISBN: 978-1-61782-980-2

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2011) by SIAM: Society for Industrial and Applied Mathematics
All rights reserved.

Printed by Curran Associates, Inc. (2011)

For permission requests, please contact SIAM: Society for Industrial and Applied Mathematics
at the address below.

SIAM
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA

Phone: (215) 382-9800
Fax: (215) 386-7999

siambooks@siam.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

VOLUME 1

SESSION S1 - ANOMALIES AND ANONYMITY

ACE: Anomaly Clustering Ensemble for Multi-Perspective Anomaly Detection in Robot Behaviors	1
<i>Shin Ando, Einoshin Suzuki, Yoichi Seki, Theerasak Thanongphongphan, Daisuke Hoshino</i>	
Interpreting and Unifying Outlier Scores	13
<i>Hans-Peter Kriegel, Peer Kroeger, Erich Schubert, Arthur Zimek</i>	
Fast Algorithms for Finding Extremal Sets	25
<i>Roberto Bayardo, Biswanath Panda</i>	
Structural Diversity for Privacy in Publishing Social Networks	35
<i>Chih-Hua Tai, Philip Yu, De-Nian Yang, Ming-Syan Chen</i>	
The Network Completion Problem: Inferring Missing Nodes and Edges in Networks	47
<i>Myunghwan Kim, Jure Leskovec</i>	

SESSION S2 - SCIENTIFIC AND MEDICAL APPLICATIONS

Integrating Distance Metrics Learned from Multiple Experts and Its Application in Inter-Patient Similarity Assessment	59
<i>Fei Wang, Jimeng Sun, Shahram Ebadollahi</i>	
Learning Feature Dependencies for Noise Correction in Biomedical Prediction	71
<i>Ghim-Eng Yap, Ah-Hwee Tan, Hwee-Hwa Pang</i>	
Data Integration via Constrained Clustering: An Application to Enzyme Clustering	83
<i>Elisa Lima, Raquel Minardi, Wagner Meira Jr., Mohammed Zaki</i>	
A Gaussian Process Based Online Change Detection Algorithm for Monitoring Periodic Time Series	95
<i>Varun Chandola, Ranga Vatsavai</i>	
Discovering Dynamic Dipoles in Climate Data, University of Minnesota	107
<i>Jaya Kawale, Michael Steinbach, Vipin Kumar</i>	

SESSION S3 - GRAPH MINING

Centralities in Large Networks: Algorithms and Observations	119
<i>U. Kang, Spiros Papadimitriou, Jimeng Sun, Hanghang Tong</i>	
Large Scale Graph Mining and Inference for Malware Detection	131
<i>Duen Horng "Polo" Chau, Carey Nachenberg, Jeffrey Wilhelm, Adam Wright, Christos Faloutsos</i>	
Non-Negative Residual Matrix Factorization with Application to Graph Anomaly Detection	143
<i>Hanghang Tong, Ching-Yung Lin</i>	
Kernel-Based Similarity Search in Massive Graph Databases with Wavelet Trees	154
<i>Yasuo Tabei, Koji Tsuda</i>	
Clustered Low Rank Approximation of Graphs in Information Science Applications	164
<i>Berkant Savas, Inderjit Dhillon</i>	

SESSION S4 – CLASSIFICATION

Who Should Label What? Instance Allocation in Multiple Expert Active Learning	176
<i>Byron Wallace, Kevin Small, Carla Brodley, Thomas Trikalinos</i>	
A Quadratic Mean based Supervised Learning Model for Managing Data Skewness	188
<i>Wei Liu, Sanjay Chawla</i>	
MKBoost: A Framework of Multiple Kernel Boosting	199
<i>Hao Xia, Steven C.H. Hoi</i>	
Efficient Kernel Approximation for Large-Scale Support Vector Machine Classification	211
<i>Ming-Syan Chen, Keng-Pei Lin</i>	
A Sequential Dual Method for Structural SVMs	223
<i>Shirish Shevade, P. Balamurugan, S. Sundararajan, S. Sathya Keerthi</i>	

SESSION S5 - FEATURE SELECTION AND DIMENSIONALITY REDUCTION

Segmented Nestedness in Binary Data	235
<i>Esa Junttila, Petteri Kaski</i>	
Extracting Interpretable Features for Early Classification on Time Series	247
<i>Zhengzheng Xing, Jian Pei, Philip Yu, Ke Wang</i>	
Feature-Based Inductive Transfer Learning through Minimum Encoding	259
<i>Shao Hao, Einoshin Suzuki</i>	
Chernoff Dimensionality Reduction - Where Fisher Meets FKT	271
<i>Jing Peng, Guna Seetharaman, Stefan Robila, Aparna Varde, Wei Fan</i>	
Weighted Rank-One Binary Matrix Factorization	283
<i>Haibing Lu, Jaideep Vaidya, Vijayalakshmi Atluri, Heechang Shin, Lili Jiang</i>	

SESSION S6 – CLUSTERING

ABACUS: Mining Arbitrary Shaped Clusters from Large Datasets based on Backbone Identification	295
<i>Vineet Chaoji, Geng Li, Hilmi Yildirim, Mohammed Zaki</i>	
Spatially-Aware Comparison and Consensus for Clusterings	307
<i>Parasaran Raman, Jeff Phillips, Suresh Venkatasubramanian</i>	
SONAR: Signal De-Mixing for Robust Correlation Clustering	319
<i>Claudia Plant</i>	
Nonparametric Bayesian Co-Clustering Ensembles	331
<i>Pu Wang, Kathryn Laskey, Carlotta Domeniconi, Michael Jordan</i>	
A Generalized Framework for Mining Arbitrarily Positioned Overlapping Co-Clusters	343
<i>Omar Odibat, Chandan K. Reddy</i>	

SESSION S7 - SOCIAL NETWORK ANALYSIS

On Node Classification in Dynamic Content-Based Networks	355
<i>Charu Aggarwal, Nan Li</i>	
Predicting Item Adoption Using Social Correlation	367
<i>Freddy Chong Tat Chua, Hady Lauw, Ee-Peng Lim</i>	
Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate	379
<i>Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui S</i>	
Towards Community Detection in Locally Heterogeneous Networks	391
<i>Charu Aggarwal, Yan Xie, Philip Yu</i>	
Cost of Collaboration vs Individual Effort in Social Networks	403
<i>Anand Santhanakrishnan, Rajarathnam Chandramouli, Kodavayur Subbalakshmi</i>	

SESSION S8 - PROBABILISTIC AND STATISTICAL METHODS

Are Your Items in Order	414
<i>Nikolaj Tatti</i>	
Probabilistic Models over Ordered Partitions with Applications in Document Ranking and Collaborative Filtering	426
<i>Truyen Tran, Dinh Phung, Svetha Venkatesh</i>	
Distributed Monitoring of the R^2 Statistic for Linear Regression	438
<i>Kanishka Bhaduri, Kamalika Das, Chris Giannella</i>	
Block-LDA: Jointly Modeling Entity-Annotated Text and Entity-Entity Links	450
<i>Ramnath Balasubramanian, William Cohen</i>	
A Tractable Pseudo-Likelihood Function for Bayes Nets Applied to Relational Data	462
<i>Oliver Schulte</i>	

SESSION S9 - TEXT AND WEB MINING

Sparse Latent Semantic Analysis	474
<i>Xi Chen, Yanjun Qi, Bing Bai, qihang Lin, Jaime Carbonell</i>	
Graph-Based Marginal Ranking for Update Summarization	486
<i>Xuan Li, Liang Du, Yi-Dong Shen</i>	

Exploiting Coherence for the Simultaneous Discovery of Latent Facets and Associated Sentiments	498
<i>Himabindu Lakkaraju, Chiranjib Bhattacharyya, Indrajit Bhattacharya, Srujana Merugu</i>	

VOLUME 2

Semi-Supervised Convolution Graph Kernels for Relation Extraction	510
<i>Xia Ning, Yanjun Qi</i>	
On Flow Authority Discovery in Social Networks	522
<i>Charu Aggarwal, Arijit Khan, Xifeng Yan</i>	

SESSION S10 - FREQUENT PATTERN MINING

Characterizing Uncertain Data using Compression	534
<i>Francesco Bonchi, Matthijs van Leeuwen, Antti Ukkonen</i>	
From Black and White to Full Colour: Extending Redescription Mining Outside the Boolean World	546
<i>Esther Galbrun, Pauli Miettinen</i>	
A Structure Function for Transaction Data	558
<i>Arno Siebes, Rene Kersten</i>	
Multidimensional Association Rules in Boolean Tensors	570
<i>Kim-Ngan Nguyen, Loic Cerf, Marc Plantevit, Jean-Francois Boulicaut</i>	
Fast Rule Mining Over Multi-Dimensional Windows	582
<i>Mahashweta Das, P. Deepak, Prasad Deshpande, Ramakrishnan Kannan</i>	

SESSION S11 - NOVEL LEARNING METHODS

FAMER: Making Multi-Instance Learning Better and Faster	594
<i>Wei Ping, Ye Xu, Jianyong Wang, Xian-Sheng Hua</i>	
Multi-Instance Mixture Models	606
<i>James Foulds, Padhraic Smyth</i>	
Multi-Label Collective Classification	618
<i>Xiangnan Kong, Xiaoxiao SHI, Philip Yu</i>	
A Probabilistic Hierarchical Approach for Pattern Discovery in Collaborative Filtering Data	630
<i>Nicola Barbieri, Giuseppe Manco, Ettore Ritacco</i>	
Online Max-Margin Weight Learning for Markov Logic Networks	642
<i>Tuyen Huynh, Raymond Mooney</i>	

SESSION S12 - DATA STREAMS AND TIME SERIES

On Classification of Graph Streams	652
<i>Charu Aggarwal</i>	
Temporal Structure Learning for Clustering Massive Data Streams in Real-Time	664
<i>Michael Hahsler, Margaret Dunham</i>	
Significance of Patterns in Time Series Collections	676
<i>Niko Vuokko, Petteri Kaski</i>	
Time Series Motifs Statistical Significance	687
<i>Nuno Castro, Paulo Azevedo</i>	
A Complexity-Invariant Distance Measure for Time Series	699
<i>Gustavo Batista, Xiaoyue Wang, Eamonn Keogh</i>	

SHORT PRESENTATIONS

On Anonymization of Multi-Graphs	711
<i>Chun Li, Charu Aggarwal, Jianyong Wang</i>	
Scalable Software-Defect Localisation by Hierarchical Mining of Dynamic Call Graphs	723
<i>Frank Eichinger, Christopher Ofner, Klemens Böhm</i>	
Computationally Generated Cardiac Biomarkers: Heart Rate Patterns to Predict Death Following Coronary Attacks	735
<i>Chih-Chun Chia, Zeeshan Syed</i>	

Unsupervised Disaggregation of Low Frequency Power Measurements	747
<i>Hyungsul Kim, Manish Marwah, Martin Arlitt, Geoff Lyon, Jiawei Han</i>	
WindMine: Fast and Effective Mining of Web-Click Sequences	759
<i>Yasushi Sakurai, Lei Li, Yasuko Matsubara, Christos Faloutsos</i>	
A Fast Algorithm for Sparse PCA and a New Sparsity Control Criteria	771
<i>Yunlong He, Renato Monteiro, Haesun Park</i>	
Gaussian Process for Dimensionality Reduction in Transfer Learning	783
<i>Bin Tong, Junbin Gao, Nguyen Huy Thach, Einoshin Suzuki</i>	
Semantic Graph Kernels for Automated Reasoning	795
<i>Evgeni Tsivtsivadze, Josef Urban, Herman Geuvers, Tom Heskes</i>	
The Odd One Out: Identifying and Characterising Anomalies	804
<i>Koen Smets, Jilles Vreeken</i>	
Sparse Solutions for Single Class SVMs: A Bi-Criterion Approach	816
<i>Santanu Das, Nikunj Oza</i>	
Multi-task Multiple Kernel Learning	828
<i>Pratik Jawanpuria, J. Saketha Nath</i>	
The Infinite Push: A New Support Vector Ranking Algorithm that Directly Optimizes Accuracy at the Absolute Top of the List	839
<i>Shivani Agarwal</i>	
Active Teaching for Inductive Learners	851
<i>Jun Du, Charles Ling</i>	
Semi-Supervised Variable Weighting for Clustering	862
<i>Ling Chen, Chengqi Zhang</i>	
Discovering Bucket Orders from Data	872
<i>Vinayaka Pandit, Sreyash Kenkre, Arindam Khan</i>	
Exemplar-Based Robust Coherent Biclustering	884
<i>Kewei Tu, Xixiu Ouyang, Dingyi Han, Vasant Honavar</i>	
Sequential Minimal Optimization in Adaptive-Bandwidth Convex Clustering	896
<i>Rikiya Takahashi</i>	
Efficient Document Clustering via Online Nonnegative Matrix Factorizations	908
<i>Fei Wang, Ping Li, Chenhao Tan, Christian König</i>	
Extending Consensus Clustering to Explore Multiple Clustering Views	920
<i>Yi Zhang, Tao Li</i>	
Weighted Graph Compression for Parameter-Free Clustering With PaCCo	932
<i>Nikola Mueller, Katrin Haegler, Junming Shao, Claudia Plant, Christian Böhm</i>	
iMet: Interactive Metric Learning in Healthcare Applications	944
<i>Fei Wang, Jimeng Sun, Jianying Hu, Shahram Ebadollahi</i>	
Maximising the Quality of Influence	956
<i>Charanpal Dhanjal, Stéphan Cléménçon</i>	
Human Dynamics in Large Communication Networks	968
<i>Pedro Olmo Vaz de Melo, Christos Faloutsos, Antonio Loureiro</i>	
Diversified Trajectory Pattern Ranking in Geo-Tagged Social Media	980
<i>Zhijun Yin, Liangliang Cao, Jiawei Han, Jiebo Luo, Thomas Huang</i>	
One-Class-Based Uncertain Data Stream Learning	992
<i>Bo Liu, Yanshan Xiao, Longbing Cao, Philip Yu</i>	
Online Discovery of Top-k Similar Motifs in Time Series Data	1004
<i>Hoang Thanh Lam, Toon Calders, Ninh Pham</i>	
Author Index	