11th SIAM International Conference on Data Mining 2011

Mesa, Arizona, USA
28-30 April 2011

Volume 1 of 2

# TABLE OF CONTENTS

## VOLUME 1

### SESSION S1 - ANOMALIES AND ANONYMITY

**ACE: Anomaly Clustering Ensemble for Multi-Perspective Anomaly Detection in Robot Behaviors** ........................................ 1  
Shin Ando, Einoshin Suzuki, Yoichi Seki, Theerasak Thanongphongphan, Daisuke Hoshino

**Interpreting and Unifying Outlier Scores** .......................................................... 13  
Hans-Peter Kriegel, Peer Kroeger, Erich Schubert, Arthur Zimek

**Fast Algorithms for Finding Extremal Sets** ......................................................... 25  
Roberto Bayardo, Biswanath Panda

**Structural Diversity for Privacy in Publishing Social Networks** .......................... 35  
Chih-Hua Tai, Philip Yu, De-Nian Yang, Ming-Syan Chen

**The Network Completion Problem: Inferring Missing Nodes and Edges in Networks**  47  
Myunghwan Kim, Jure Leskovec

### SESSION S2 - SCIENTIFIC AND MEDICAL APPLICATIONS

**Integrating Distance Metrics Learned from Multiple Experts and Its Application in Inter-Patient Similarity Assessment** ................................................................. 59  
Fei Wang, Jimeng Sun, Shahram Ebadollahi

**Learning Feature Dependencies for Noise Correction in Biomedical Prediction** ................................................................. 71  
Ghim-Eng Yap, Ah-Hwee Tan, Hwee-Hwa Pang

**Data Integration via Constrained Clustering: An Application to Enzyme Clustering** ................................................................. 83  
Elisa Lima, Raquel Minardi, Wagner Meira Jr., Mohammed Zaki

**A Gaussian Process Based Online Change Detection Algorithm for Monitoring Periodic Time Series** ......................................................... 95  
Varun Chandola, Ranga Vatsavai

**Discovering Dynamic Dipoles in Climate Data, University of Minnesota** ................. 107  
Jaya Kawale, Michael Steinbach, Vipin Kumar

### SESSION S3 - GRAPH MINING

**Centralities in Large Networks: Algorithms and Observations** ................................. 119  
U. Kang, Spiros Papadimitriou, Jimeng Sun, Hanghang Tong

**Large Scale Graph Mining and Inference for Malware Detection** .......................... 131  
Duen Horng “Polo” Chau, Carey Naienberger, Jeffrey Wilhelm, Adam Wright, Christos Faloutsos

**Non-Negative Residual Matrix Factorization with Application to Graph Anomaly Detection** ................................................................. 143  
Hanghang Tong, Ching-Yung Lin

**Kernel-Based Similarity Search in Massive Graph Databases with Wavelet Trees** ................................. 154  
Yasuo Tabei, Koji Tsuda

**Clustered Low Rank Approximation of Graphs in Information Science Applications** ................................................................. 164  
Berkant Savas, Inderjit Dhillon

### SESSION S4 – CLASSIFICATION

**Who Should Label What? Instance Allocation in Multiple Expert Active Learning** ................................. 176  
Byron Wallace, Kevin Small, Carla Brodley, Thomas Trikalinos

**A Quadratic Mean based Supervised Learning Model for Managing Data Skewness** ................................................................. 188  
Wei Liu, Sanjay Chawla

**MKBoost: A Framework of Multiple Kernel Boosting** ............................................. 199  
Hao Xia, Steven C.H. Hoi

**Efficient Kernel Approximation for Large-Scale Support Vector Machine Classification** ................................................................. 211  
Ming-Syan Chen, Keng-Pei Lin

**A Sequential Dual Method for Structural SVMs** .................................................... 223  
Shirish Shevade, P. Balaraman, S. Sundararajan, S. Sathiya Keerthi
SESSION S5 - FEATURE SELECTION AND DIMENSIONALITY REDUCTION

Segmented Nestedness in Binary Data ................................................................. 235
Esa Junttila, Petteri Kaski

Extracting Interpretable Features for Early Classification on Time Series ........................................ 247
Zhengzheng Xing, Jian Pei, Philip Yu, Ke Wang

Feature-Based Inductive Transfer Learning through Minimum Encoding ........................................ 259
Shao Hao, Einoshin Suzuki

Chernoff Dimensionality Reduction - Where Fisher Meets FKT ...................................................... 271
Jing Peng, Guna Seetharaman, Stefan Robila, Aparna Varde, Wei Fan

Weighted Rank-One Binary Matrix Factorization ................................................................. 283
Haibing Lu, Jaideep Vaidya, Vijayalakshmi Alluri, Heechang Shin, Lili Jiang

SESSION S6 – CLUSTERING

ABACUS: Mining Arbitrary Shaped Clusters from Large Datasets based on Backbone Identification ........ 295
Vineet Chaoji, Geng Li, Hilmi Yildirim, Mohammed Zaki

Spatially-Aware Comparison and Consensus for Clusterings ......................................................... 307
Parasaran Raman, Jeff Phillips, Suresh Venkatasubramanian

SONAR: Signal De-Mixing for Robust Correlation Clustering ......................................................... 319
Claudia Plant

Nonparametric Bayesian Co-Clustering Ensembles ................................................................. 331
Pu Wang, Kathryn Laskey, Carlotta Domeniconi, Michael Jordan

A Generalized Framework for Mining Arbitrarily Positioned Overlapping Co-Clusters ................ 343
Omar Odibat, Chandan K. Reddy

SESSION S7 - SOCIAL NETWORK ANALYSIS

On Node Classification in Dynamic Content-Based Networks ......................................................... 355
Charu Aggarwal, Nan Li

Predicting Item Adoption Using Social Correlation ................................................................. 367
Freddy Chong Tat Chua, Hady Laoui, Ee-Peng Lim

Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate ..................................................................................... 379
Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui S

Towards Community Detection in Locally Heterogeneous Networks ........................................ 391
Charu Aggarwal, Yan Xie, Philip Yu

Cost of Collaboration vs Individual Effort in Social Networks ......................................................... 403
Anand Santhanakrishnan, Rajarathnam Chandramouli, Kodavayur Subbalakshmi

SESSION S8 - PROBABILISTIC AND STATISTICAL METHODS

Are Your Items in Order ................................................................................................. 414
Nikolaj Tatti

Probabilistic Models over Ordered Partitions with Applications in Document Ranking and Collaborative Filtering ........................................................................ 426
Truyen Tran, Dinh Phung, Svetla Venkatesh

Distributed Monitoring of the $R^2$ Statistic for Linear Regression ........................................ 438
Kanishka Bhaduri, Kamalika Das, Chris Giannella

Block-LDA: Jointly Modeling Entity-Annotated Text and Entity-Entity Links ........................ 450
Ramnath Balasubramanian, William Cohen

A Tractable Pseudo-Likelihood Function for Bayes Nets Applied to Relational Data ................ 462
Oliver Schulte

SESSION S9 - TEXT AND WEB MINING

Sparse Latent Semantic Analysis ................................................................................................. 474
Xi Chen, Yanjun Qi, Bing Bai, qihang Lin, Jaime Carbonell

Graph-Based Marginal Ranking for Update Summarization ......................................................... 486
Xuan Li, Liang Du, Yi-Dong Shen
Exploiting Coherence for the Simultaneous Discovery of Latent Facets and Associated Sentiments ........................................... 498
Himabindu Lakkaraju, Chiranjib Bhattacharyya, Indrajit Bhattacharya, Srujana Merugu

VOLUME 2

Semi-Supervised Convolution Graph Kernels for Relation Extraction ..................................................................................... 510
Xia Ning, Yanjun Qi

On Flow Authority Discovery in Social Networks .................................................... 522
Charu Aggarwal, Arijit Khan, Xifeng Yan

SESSION S10 - FREQUENT PATTERN MINING

Characterizing Uncertain Data using Compression .......................................................... 534
Francesco Bonchi, Matthijs van Leeuwen, Antti Ukkonen

From Black and White to Full Colour: Extending Redescription Mining Outside the Boolean World ........................................... 546
Esther Galbrun, Pauli Miettinen

A Structure Function for Transaction Data ................................................................ 558
Arno Siebes, Rene Kersten

Multidimensional Association Rules in Boolean Tensors .............................................. 570
Kim-Ngan Nguyen, Loic Cerf, Marc Plantevit, Jean-Francois Boulicaut

Fast Rule Mining Over Multi-Dimensional Windows ................................................ 582
Mahashweta Das, P. Deepak, Prasad Deshpande, Ramakrishnan Kannan

SESSION S11 - NOVEL LEARNING METHODS

FAMER: Making Multi-Instance Learning Better and Faster ...................................... 594
Wei Ping, Ye Xu, Jianyong Wang, Xian-Sheng Hua

Multi-Instance Mixture Models ................................................................................ 606
James Foulds, Padhraic Smyth

Multi-Label Collective Classification ......................................................................... 618
Xiangnan Kong, Xiaoxiao Shi, Philip Yu

A Probabilistic Hierarchical Approach for Pattern Discovery in Collaborative Filtering Data ......................................................... 630
Nicola Barbieri, Giuseppe Manco, Ettore Ritacco

Online Max-Margin Weight Learning for Markov Logic Networks .......................... 642
Tuyen Huynh, Raymond Mooney

SESSION S12 - DATA STREAMS AND TIME SERIES

On Classification of Graph Streams ........................................................................... 652
Charu Aggarwal

Temporal Structure Learning for Clustering Massive Data Streams in Real-Time ....................... 664
Michael Hahsler, Margaret Dunham

Significance of Patterns in Time Series Collections ................................................... 676
Niko Vuokko, Petteri Kaski

Time Series Motifs Statistical Significance .................................................................. 687
Nuno Castro, Paulo Azevedo

A Complexity-Invariant Distance Measure for Time Series ......................................... 699
Gustavo Batista, Xiaoqyue Wang, Eamonn Keogh

SHORT PRESENTATIONS

On Anonymization of Multi-Graphs ............................................................................. 711
Chun Li, Charu Aggarwal, Jianyong Wang

Scalable Software-Defect Localisation by Hierarchical Mining of Dynamic Call Graphs ............................................................. 723
Frank Eichinger, Christopher Offner, Klemens Böhm

Computationally Generated Cardiac Biomarkers: Heart Rate Patterns to Predict Death Following Coronary Attacks .................. 735
Chih-Chun Chia, Zeeshan Syed