

SIAM International Conference on Data Mining 2014

Philadelphia, Pennsylvania, USA
24 – 26 April 2014

Volume 1 of 2

Editors:

**Mohammed Zaki
Zoran Obradovic
Pang Ning Tan**

**Arindam Banerjee
Chandrika Kamath
Srinivasan Parthasarathy**

ISBN: 978-1-5108-1151-5

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© (2014) by SIAM: Society for Industrial and Applied Mathematics
All rights reserved.

Printed by Curran Associates, Inc. (2015)

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

Session I: Classification 1

Class Augmented Active Learning	1
Hong Cao, Chunyu Bao, Xiao-Li Li, Yew-Kwong Woon	
Batch Mode Active Learning with Hierarchical-Structured Embedded Variance	10
Yu Cheng, Zhengzhang Chen, Hongliang Fei, Fei Wang, Alok Choudhary	
Learning from Multi-User Multi-Attribute Annotations	19
Ou Wu, Shuxiao Li, Honghui Dong, Ying Chen, Weiming Hu	
Selective Sampling on Probabilistic Data	28
Peng PengG, Raymond Chi-Wing Wong	
Disambiguation-Free Partial Label Learning	37
Min-Ling Zhang	

Session 2: Networks, Graphs 1

DAVA: Distributing Vaccines over Networks under Prior Information.....	46
Yao Zhang, B. Aditya Prakash	
Influence Maximization with Viral Product Design	55
Nicola Barbieri, Francesco Bonchi	
A Generative Model with Network Regularization for Semi-Supervised Collective Classification	64
Ruichao Shi, Qingyao Wu, Yuming Ye, Shen-Shyang Ho	
Local Learning for Mining Outlier Subgraphs from Network Datasets	73
Manish Gupta, Arun Mallya, Subhro Roy, Jason H.D. Cho, Jiawei Han	
A Probabilistic Approach to Uncovering Attributed Graph Anomalies	82
Nan Li, Huan Sun, Kyle Chipman, Jemin George, Xifeng Yan	

Session 3: Large Scale, Matrix/Tensor

VoG: Summarizing and Understanding Large Graphs	91
Danai Koutra, U Kang, Jilles Vreeken, Christos Faloutsos	
Context-Preserving Hashing for Fast Text Classification.....	100
Lianhua Chi, Bin Li, Xingquan Zhu	

FlexiFaCT: Scalable Flexible Factorization of Coupled Tensors on Hadoop	109
Alex Beutel, Abhimanu Kumar, Evangelos Papalexakis, Partha Pratim, Christos Faloutsos, Eric Xing	
Turbo-SMT: Accelerating Coupled Sparse Matrix-Tensor Factorizations by 200x	118
Evangelos Papalexakis, Tom Mitchell, Nicholas Sidiropoulos, Christos Faloutsos, Partha Talukdar, Brian Murphy	
DuSK: A Dual Structure-preserving Kernel for Supervised Tensor Learning with Applications to Neuroimages	127
Lifang He, Xiangnan Kong, Philip Yu, Ann Ragin, Zhifeng Hao, Xiaowei Yang	

Session 4: Feature Selection/Extraction

How Can I Index My Thousands of Photos Effectively and Automatically? An Unsupervised Feature Selection Approach.....	136
Juhua Hu, Jian Pei, Jie Tang	
Unveiling Variables in Systems of Linear Equations.....	145
Behzad Golshan, Evimaria Terzi	
Transductive HSIC Lasso	154
Dan He, Irina Rish, Laxmi Parida	
Robust Subspace Discovery through Supervised Low-Rank Constraints.....	163
Sheng Li, Yun Fu	
Adaptive Quantization for Hashing: An Information-Based Approach to Learning Binary Codes	172
Caiming Xiong, Wei Chen, Gang Chen, David Johnson, Jason Corso	

Session 5: Multi-task/Transfer Learning

Linking Heterogeneous Input Spaces with Pivots for Multi-Task Learning.....	181
Jingrui He, Yan Liu, Qiang Yang	
Active Multitask Learning Using Both Latent and Supervised Shared Topics	190
Ayan Acharya, Raymond Mooney, Joydeep Ghosh	
Multi-Task Feature Selection on Multiple Networks via Maximum Flows.....	199
Mahito Sugiyama, Chloé-Agathe Azencott, Dominik Grimm, Yoshinobu Kawahara, Karsten Borgwardt	
Mixed-Transfer: Transfer Learning over Mixed Graphs.....	208
Ben Tan, Erheng Zhong, Michael Ng, Qiang Yang	
Multi-Graph Learning with Positive and Unlabeled Bags.....	217
Jia Wu, Zhibin Hong, Shirui Pan, Xingquan Zhu, Chengqi Zhang, Zhihua Cai	

Session 6: Applications

Consumer Segmentation and Knowledge Extraction from Smart Meter and Survey Data	226
Tri Kurniawan Wijaya, Tanuja Ganu, Dipanjan Chakraborty, Karl Aberer, Deva Seetharam	
Keeping up with Innovation: A Predictive Framework for Modeling Healthcare Data with Evolving Clinical Interventions.....	235
Sunil Gupta, Santu Rana, Dinh Phung, Svetha Venkatesh	
Turning the Tide: Curbing Deceptive Yelp Behaviors.....	244
Mahmudur Rahman, Bogdan Carbutar, Jaime Ballesteros, George Burri, Duen Horng Chau	
Predictive Learning in the Presence Of Heterogeneity and Limited Training Data	253
Anuj Karpatne, Ankush Khandelwal, Shyam Boriah, Vipin Kumar	
Forecasting a Moving Target: Ensemble Models for ILI Case Count Predictions.....	262
Prithwish Chakraborty, Pejman Khadivi, Bryan Lewis, Aravindan Mahendiran, Jiangzhuo Chen, Patrick Butler, Elaine Nsoesie, Sumiko Mekaru, John Brownstein, Madhav Marathe, Naren Ramakrishnan	

Session 7: Classification 2

AUC Dominant Unsupervised Ensemble of Binary Classifiers	271
Priyanka Agrawal, Vikas Raykar, Amrita Saha	
Convex Optimization for Binary Classifier Aggregation in Multiclass Problems	280
Sunho Park, Tae Hyun Hwang, Seungjin Choi	
A Deep Learning Approach to Link Prediction in Dynamic Networks.....	289
Xiaoyi Li, Nan Du, Hui Li, Kang Li, Jing Gao, Aidong Zhang	
A Marginalized Denoising Method for Link Prediction in Relational Data.....	298
Zheng Chen, Weixiong Zhang	
Learning on Probabilistic Labels.....	307
Peng Peng, Raymond Chi-Wing Wong, Philip Yu	

Session 8: Applications

Adaptive User Distance Modeling in Social Media.....	316
Erheng Zhong, Wei Fan, Qiang Yang	
Make It or Break It: Manipulating Robustness in Large Networks	325
Hau Chan, Leman Akoglu, Hanghang Tong	

It Takes Two to Tango: Modeling Social Tie Development with Both Online and Offline Interactions..... 334
Peifeng Yin, Qi He, Xingjie Liu, Wang-Chien Lee

Laplacian Spectral Properties of Graphs from Random Local Samples..... 343
Zhengwei Wu, Victor Preciado

Triangle Counting in Streamed Graphs Via Small Vertex Covers 352
David García-Soriano, Konstantin Kutzkov

Session 9: Probabilistic Topic Modeling

A Constrained Hidden Markov Model Approach for Non-Explicit Citation Context Extraction 361
Parikshit Sondhi, ChengXiang Zhai

Joint Author Sentiment Topic Model..... 370
Subhabrata Mukherjee, Gaurab Basu, Sachindra Joshi

A Dynamic Nonparametric Model for Characterizing the Topical Communities in Social Streams 379
Ziqi Liu, Qinghua Zheng, Fei Wang, Zhenhua Tian, Bo Li

Recurrent Chinese Restaurant Process with a Duration-based Discount for Event Identification from Twitter 388
Qiming Diao, Jing Jiang

Automatic Construction and Ranking of Topical Keyphrases on Collections of Short Documents 398
Marina Danilevsky, Chi Wang, Nihit Desai, Xiang Ren, Jingyi Guo, Jiawei Han

Session 10: Clustering and Density Estimation

On Randomly Projected Hierarchical Clustering with Guarantees 407
Michail Vlachos

Self-Taught Spectral Clustering via Constraint Augmentation 416
Xiang Wang, Jun Wang, Buyue Qian, Fei Wang, Ian Davidson

Embed and Conquer: Scalable Embeddings for Kernel k-Means on MapReduce..... 425
Ahmed Elgohary, Ahmed Farahat, Mohamed Kamel, Fakhri Karray

A Constructive Setting for the Problem of Density Ratio Estimation..... 434
Vladimir Vapnik, Igor Braga, Rauf Izmailov

Density Estimation with Adaptive Sparse Grids for Large Data Sets 443
Benjamin Peherstorfer, Dirk Pflüger, Hans-Joachim Bungartz

Session 11: Recommendation and Social Media Analysis

Latent Factor Transition for Dynamic Collaborative Filtering	452
Chenyi Zhang, Ke Wang, Hongkun Yu, Jianling Sun, Ee-Peng Lim	
Contextual Combinatorial Bandit and its Application on Diversified Online Recommendation.....	461
Lijing Qin, Shouyuan Chen, Xiaoyan Zhu	
User Preference Learning with Multiple Information Fusion for Restaurant Recommendation.....	470
Yanjie Fu, Bin Liu, Yong Ge, Zijun Yao, Hui Xiong	
On Modeling Community Behaviors and Sentiments in Microblogging	479
Tuan-Anh Hoang, William W. Cohen, Ee-Peng Lim	
Selecting a Representative Set of Diverse Quality Reviews Automatically	488
Nana Xu, Hongyan Liu, Jiawei Chen, Jun He, Xiaoyong Du	

Session 12: Pattern mining and Time Series

Finding the True Frequent Itemsets.....	497
Matteo Riondato, Fabio Vandin	
A Statistical Learning Theory Framework for Supervised Pattern Discovery	506
Jonathan Huggins, Cynthia Rudin	
When and Where: Predicting Human Movements Based on Social Spatial-Temporal Events.....	515
Ning Yang, Xiangnan Kong, Fengjiao Wang, Philip Yu	
Ensembles of Elastic Distance Measures for Time Series Classification	524
Jason Lines, Anthony Bagnall	
Discovery of Rare Sequential Topic Patterns in Document Stream.....	533
Zhongyi Hu, Hongan Wang, Jiaqi Zhu, Maozhen Li, Ying Qiao, Changzhi Deng	

Posters

Generalized Outlier Detection with Flexible Kernel Density Estimates	542
Erich Schubert, Arthur Zimek, Hans-Peter Kriegel	
Building Associations between Markers of Environmental Stressors and Adverse Human Health Impacts Using Frequent Itemset Mining.....	551
Shannon Bell, Stephen Edwards	
Less is More: Similarity of Time Series under Linear Transformations.....	560
Frank Hoppner	

Graphical Models for Identifying Fraud and Waste in Healthcare Claims	569
Peder Olsen, Ramesh Natarajan, Sholom Weiss	
Classifier-Adjusted Density Estimation for Anomaly Detection and One-Class Classification	578
Lisa Friedland, Amanda Gentzel, David Jensen	
Towards Accurate Histogram Publication under Differential Privacy	587
Xiaojian Zhang, Rui Chen, Jianliang Xu, Xiaofeng Meng, Yingtao Xie	
WCAMiner: A Novel Knowledge Discovery System for Mining Concept Associations using Wikipedia.....	596
Peng Yan, Wei Jin	
META: Multi-resolution Framework for Event Summarization.....	605
Yexi Jiang, Chang-Shing Perng, Tao Li	
The Benefits of Personalized Smartphone-Based Activity Recognition Models	614
Jeffrey Lockhart, Gary Weiss	
Online Matrix Completion Through Nuclear Norm Regularisation	623
Charanpal Dhanjal, Romaric Gaudel, Stéphan Cléménçon	
Online discovery of group level contextual events in time series.....	632
Xi Chen, Abdullah Mueen, Vijay Narayanan, Nikos Karampatziakis, Gagan Bansal, Vipin Kumar	
Beating Human Analysts in Nowcasting Corporate Earnings by using Publicly Available Stock Price and Correlation Features.....	641
Michael Kamp, Mario Boley, Thomas Gärtner	
Relational Regularization and Feature Ranking	650
Fabrizio Costa, Mathias Verbeke, Luc De Raedt	
Exploiting Monotonicity Constraints in Active Learning for Ordinal Classification	659
Pieter Soons, Ad Feelders	
Quantifying Trust Dynamics in Signed Graphs, the S-Cores Approach.....	668
Christos Giatsidis, Michalis Vazirgiannis, Dimitrios Thilikos, Bogdan Cautis, Silviu Maniu	
Improving Credit Card Fraud Detection by using Calibrated Probabilities	677
Alejandro Correa Bahnsen, Djamila Aouada, Aleksandar Stojanovic, Björn Ottersten	
Unsupervised Selection of Robust Audio Feature Subsets.....	686
Gerhard Sageder, Maia Zaharieva, Matthias Zeppelzauer	
Kaleido: Network Traffic Attribution using Multifaceted Footprinting	695
Ting Wang, Fei Wang, Reiner Sailer, Douglas Schales	

ROCsearch - An ROC-guided Search Strategy for Subgroup Discovery.....	704
Marvin Meeng, Wouter Duivesteijn, Arno Knobbe	
Result Integrity Verification of Outsourced Bayesian Network Structure Learning.....	713
Ruilin Liu, Hui (Wendy) Wang, Changhe Yuan	
Classifying Imbalanced Data Streams via Dynamic Feature Group Weighting with Importance Sampling	722
Ke Wu, Andrea Edwards, Wei Fan, Jing Gao, Kun Zhang	
Towards Breaking the Curse of Dimensionality for High-Dimensional Privacy	731
Hessam Zakerzadeh, Charu Aggarwal, Ken Barker	
Extracting Researcher Metadata with Labeled Features.....	740
Sujatha Das Gollapalli, Yanjun Qi, Prasenjit Mitra, C. Lee Giles	
Future Influence Ranking of Scientific Literature	749
Senzhang Wang, Sihong Xie, Xiaoming Zhang, Zhoujun Li, Philip Yu, Xinyu Shu	
Membership Detection Using Cooperative Data Mining Algorithms	758
Calvin Newport, Lisa Singh, Yiqing Ren	
Efficient Matching of Substrings in Uncertain Sequences	767
Yuxuan Li, James Bailey, Lars Kulik, Jian Pei	
Modeling Asymmetry and Tail Dependence among Multiple Variables by Using Partial Regular Vine	776
Wei Wei, Junfu Yin, Jinyan Li, Longbing Cao	
Multi-Task Clustering using Constrained Symmetric Non-Negative Matrix Factorization.....	785
Samir Al-Stouhi, Chandan Reddy	
A Weighted Adaptive Mean Shift Clustering Algorithm	794
Yazhou Ren, Carlotta Domeniconi, Guoji Zhang, Guoxian Yu	
Community Discovery in Social Networks via Heterogeneous Link Association and Fusion.....	803
Lei Meng, Ah-Hwee Tan	
Large-scale Kernel RankSVM.....	812
Tzu-Ming Kuo, Ching-Pei Lee, Chih-Jen Lin	
Learning Influence Relationships from Graphs.....	821
Chuan Hu, Huiping Cao, Chaomin Ke	
Discriminative Density-ratio Estimation	830
Yun-Qian Miao, Ahmed Farahat, Mohamed Kamel	

Density-Based Clustering Validation	839
Davoud Moulavi, Pablo Jaskowiak, Ricardo Campello, Arthur Zimek, Jorg Sander	
Directed Interpretable Discovery in Tensors with Sparse Projection.....	848
Chia-Tung Kuo, Ian Davidson	
Recovering Missing Labels of Crowdsourcing Workers.....	857
Qingyang Hu, Kevin Chiew, Hao Huang, Qinming He	
Factor Matrix Trace Norm Minimization for Low-Rank Tensor Completion	866
Yuanyuan Liu, Fanhua Shang, Hong Cheng, James Cheng, Hanghang Tong	
A New Framework for Traffic Anomaly Detection	875
Jinsong Lan, Cheng Long, Raymond Chi-Wing Wong, Youyang Chen, Yanjie Fu, Danhuai Guo, Shuguang Liu, Yong Ge, Yuanchun Zhou, Jianhui Li	
Characterising Seismic Data	884
Roel Bertens, Arno Siebes	
Submodularity in Team Formation Problem.....	893
Avrdeep Bhowmik, Vivek Borkar, Dinesh Garg, Madhavan pallan	
Unsupervised Feature Learning by Deep Sparse Coding	902
Yunlong He, Koray Kavukcuoglu, Yun Wang, Arthur Szalm, Yanjun Qi	
Online Anomaly Detection by Improved Grammar Compression of Log Sequences	911
Yun Gao, Wei Zhou, Zhang Zhang, Jizhong Han, Dan Meng, Zhiyong Xu	
Large-Scale Multi-Label Learning with Incomplete Label Assignments.....	920
Xiangnan Kong, Zhaoming Wu, Li-Jia Li, Ruofei Zhang, Philip Yu, Hang Wu, Wei Fan	
Adversarial Learning with Bayesian Hierarchical Mixtures of Experts	929
Yan Zhou, Murat Kantarcioglu	
Discriminant Analysis for Unsupervised Feature Selection.....	938
Jiliang Tang, Xia Hu, Huiji Gao, Huan Liu	
Finding Friends on a New Site Using Minimum Information	947
Reza Zafarani, Huan Liu	
On Finding the Point Where There Is No Return: Turning Point Mining on Game Data	956
Wei Gong, Ee-Peng Lim, Feida Zhu, Palakorn Achananuparp, David Lo	
Addressing Human Subjectivity via Transfer Learning: An Application to Predicting Disease Outcome in Multiple Sclerosis Patients	965
Yijun Zhao, Carla E. Brodley, Tanuja Chitnis, Brian C. Healy	

A Graphical Model Approach to ATLAS-Free Mining of MRI Images.....	974
Chris Magnano, Ameet Soni, Sriraam Natarajan, Gautam Kunapuli	
To Sample or To Smash? Estimating reachability in large time-varying graphs.....	983
Prithwish Basu, Feng Yu, Amotz Bar-Noy, Dror Rawitz	
An Optimization-based Framework to Learn Conditional Random Fields for Multi-label Classification	992
Mahdi Pakdaman Naeini, Iyad Batal, Zitao Liu, Charmgil Hong, Milos Hauskrecht	
Discovering Groups of Time Series with Similar Behavior in Multiple Small Intervals of Time.....	1001
Gowtham Atluri, Michael Steinbach, Kelvin Lim, Angus MacDonald III, Vipin Kumar	
Memory-efficient Query-driven Community Detection with Application to Complex Disease Associations.....	1010
Steve Harenberg, Ramona Seay, Stephen Ranshous, Kanchana Padmanabhan, Jitendra Harlalka, Eric Schendel, Michael O'Brien, Rada Chirkova, William Hendrix, Nagiza Samatova	
Frugal Traffic Monitoring with Autonomous Participatory Sensing	1019
Vladimir Coric, Nemanja Djuric, Slobodan Vucetic	
Iterative Framework Radiation Hybrid Mapping.....	1028
Raed Seetan, Anne Denton, Omar Al-Azzam, Ajay Kumar, Jazarai Sturdivant, Shahryar Kianian	
A Guide to Selecting a Network Similarity Method	1037
Sucheta Soundarajan, Tina Eliassi-Rad, Brian Gallagher	
Subgraph Search in Large Graphs with Result Diversification	1046
Huiwen Yu, Dayu Yuan	
Mining Interpretable Biomarkers from Electronic Health Record from Aftercare Home Health Providers	1055
Sanjoy Dey	
Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction	1064
Vignesh Veppur Sankaranarayanan, Junaed Sattar, Laks V.S.Lakshmanan	
Accelerating Graph Adjacency Matrix Multiplications with Adjacency Forest.....	1073
Masaaki Nishino, Norihito Yasuda, Shin-ichi Minato, Masaaki Nagata	