

SIAM International Conference on Data Mining 2013

**Austin, Texas, USA
2-4 May 2013**

ISBN: 978-1-62748-724-5

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© (2013) by SIAM: Society for Industrial and Applied Mathematics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

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 1: MINING WITH DIG DATA

| | |
|--|----|
| Fast Exact Max-Kernel Search | 1 |
| <i>Ryan Curtin, Alexander Gray, Parikshit Ram</i> | |
| Triadic Measures on Graphs: The Power of Wedge Sampling | 10 |
| <i>Tamara Kolda, Ali Pinar, C. Seshadhri</i> | |
| Maximal Deviations of Incomplete U-statistics with Applications to Empirical Risk Sampling | 19 |
| <i>Stéphan Cléménçon, Sylvain Robbiano, Jessica Tressou</i> | |
| NetSpot: Spotting Significant Anomalous Regions on Dynamic Networks | 28 |
| <i>Petko Bogdanov, Christos Faloutsos, Misael Mongiovi, Evangelos Papalexakis, Razvan Ranca, Ambuj Singh</i> | |
| Mining Connection Pathways for Marked Nodes in Large Graphs | 37 |
| <i>Leman Akoglu, Duen Horng Chau, Christos Faloutsos, Nikolaj Tatti, Hanghang Tong, Jilles Vreeken</i> | |

SESSION 2: MINING WITH UNCERTAIN AND NOISY DATA

| | |
|---|----|
| Missing or Inapplicable: Treatment of Incomplete Continuous-valued Features in Supervised Learning | 46 |
| <i>Lei Liu, Prakash Mandayam Comar, Antonio Nucci, Sabyasachi Saha, Pang-Ning Tan</i> | |
| Patient Risk Prediction Model via Top-k Stability Selection | 55 |
| <i>Jianying Hu, Yashu Liu, Jimeng Sun, Jieping Ye, Jiayu Zhou</i> | |
| Collective Kernel Construction in Noisy Environment | 64 |
| <i>Chris Ding, Deguang Kong, Miao Zhang</i> | |
| Mining Probabilistic Representative Frequent Patterns From Uncertain Data | 73 |
| <i>Ling Chen, Chunyang Liu, Chengqi Zhang</i> | |
| Discriminative Feature Selection for Uncertain Graph Classification | 82 |
| <i>Xiangnan Kong, Ann Ragin, Xue Wang, Philip Yu</i> | |

SESSION 3: CLUSTERING

| | |
|--|-----|
| Determining the Number of Clusters via Iterative Consensus Clustering | 94 |
| <i>Carl Meyer, Shaina Race, Kevin Valakuzhy</i> | |
| Constrained Spectral Clustering using L1 Regularization | 103 |
| <i>Daniel Boley, Jaya Kawale</i> | |
| Efficient Anytime Density-based Clustering | 112 |
| <i>Christian Böhm, Jing Feng, Xiao He, Son T. Mai</i> | |
| Evolutionary Soft Co-Clustering | 121 |
| <i>Shuiwang Ji, Wenlu Zhang, Rui Zhang</i> | |
| Sparse Subspace Clustering via Group Sparse Coding | 130 |
| <i>Duc Son Pham, Dinh Phung, Budhaditya Saha, Svetha Venkatesh</i> | |

SESSION 4: GRAPH STREAM DATA MINING

| | |
|--|-----|
| On Graph Stream Clustering with Side Information | 139 |
| <i>Philip Yu, Yuchen Zhao</i> | |
| Dynamic Community Detection in Weighted Graph Streams | 151 |
| <i>Jian-Huang Lai, Chang-Dong Wang, Philip Yu</i> | |
| DELTA CON: A Principled Massive-Graph Similarity Function | 162 |
| <i>Christos Faloutsos, Danai Koutra, Joshua Vogelstein</i> | |
| What's Your Next Move: User Activity Prediction in Location-based Social Networks | 171 |
| <i>Hong Cheng, Jihang Ye, Zhe Zhu</i> | |
| CoFiSet: Collaborative Filtering via Learning Pairwise Preferences over Item-sets | 180 |
| <i>Li Chen, Weike Pan</i> | |

SESSION 5: ANOMALY & OUTLIER DETECTION

| | |
|--|-----|
| k-means--: A Unified Approach to Clustering and Outlier Detection | 189 |
| <i>Sanjay Chawla, Aristides Gionis</i> | |
| CMI: An Information-Theoretic Contrast Measure for Enhancing Subspace Cluster and Outlier Detection | 198 |
| <i>Klemens Böhm, Fabian Keller, Emmanuel Müller, Hoang Vu Nguyen, Jilles Vreeken</i> | |
| Cost-Sensitive Double Updating Online Learning and Its Application to Online Anomaly Detection | 207 |
| <i>Steven C.H. Hoi, Peilin Zhao</i> | |
| Efficient Selection of Globally Optimal Rules on Large Imbalanced Data Based on Rule Coverage Relationship Analysis | 216 |
| <i>Longbing Cao, Jinju Li, Can Wang, Philip S. Yu</i> | |
| Outlier Detection with Space Transformation and Spectral Analysis | 225 |
| <i>Ira Assent, Xuan Hong Dang, Barbora Micenkova, Raymond Ng</i> | |

SESSION 6: MULTI-VIEW AND MULTI-SOURCE DATA MINING

| | |
|--|-----|
| Multi-objective Multi-view Spectral Clustering via Pareto Optimization | 234 |
| <i>Ian Davidson, Buyue Qian, Xiang Wang, Jieping Ye</i> | |
| Multi-Transfer: Transfer Learning with Multiple Views and Multiple Sources | 243 |
| <i>Ben Tan, Evan Wei Xiang, Qiang Yang, Erheng Zhong</i> | |
| Multi-View Clustering via Joint Nonnegative Matrix Factorization | 252 |
| <i>Jing Gao, Jiawei Han, Jialu Liu, Chi Wang</i> | |
| On Handling Negative Transfer and Imbalanced Distributions in Multiple Source Transfer Learning | 261 |
| <i>Jing Gao, Liang Ge, Kang Li, Hung Ngo, Aidong Zhang</i> | |
| Unsupervised Feature Selection for Multi-View Data in Social Media | 270 |
| <i>Huiji Gao, Xia Hu, Huan Liu, Jiliang Tang</i> | |

SESSION 7: SEMI-SUPERVISED AND ACTIVE LEARNING

| | |
|--|-----|
| SMART: Semi-Supervised Music Emotion Recognition with Social Tagging | 279 |
| <i>Andrew Horner, Derek Hao Hu, Bin Wu, Qiang Yang, Erheng Zhong</i> | |
| Probabilistic Combination of Classifier and Cluster Ensembles for Non-transductive Learning | 288 |
| <i>Ayan Acharya, Joydeep Ghosh, Eduardo Hruschka, Jean-David Ruvini, Badrul Sarwar</i> | |
| Active Learning to Rank using Pairwise Supervision | 297 |
| <i>Ian Davidson, Hongfei Li, Buyue Qian, Jun Wang, Xiang Wang</i> | |
| ActNeT: Active Learning for Networked Texts in Microblogging | 306 |
| <i>Huiji Gao, Xia Hu, Huan Liu, Jiliang Tang</i> | |
| Active Class Discovery and Learning for Networked Data | 315 |
| <i>Meng Fang, Jie Yin, Chengqi Zhang, Xingquan Zhu</i> | |

SESSION 8: APPLICATIONS

| | |
|--|-----|
| Climate Multi-model Regression Using Spatial Smoothing | 324 |
| <i>Arindam Banerjee, Karthik Subbian</i> | |
| Distribution Regularized Regression Framework for Climate Modeling | 333 |
| <i>Zubin Abraham, Malgorzata Liszewska, Perdinan Perdinan, Pang-Ning Tan, Julie Winkler, Shiyuan Zhong</i> | |
| Sparse Representation for HIV-1 Protease Drug Resistance Prediction | 342 |
| <i>Robert Harrison, Irene Weber, Xiaxia Yu</i> | |
| Dynamic Shaker Detection from Evolving Entities | 350 |
| <i>Wei Fan, Xiaoxiao Shi, Philip Yu</i> | |
| Monitoring and mining GPS traces in transit space | 359 |
| <i>Leon Stenneth, Philip S Yu</i> | |

SESSION 9: SOCIAL NETWORK ANALYSIS

| | |
|--|-----|
| Exploiting Synchronicity Networks for Finding Valuables in Heterogeneous Networks | 369 |
| <i>Ching-Yung Lin, Zhen Wen</i> | |

| | |
|--|-----|
| Exploring and Inferring User-User Pseudo-Friendship for Sentiment Analysis with Heterogeneous Networks..... | 378 |
| <i>Hongbo Deng, Jiawei Han, Heng Ji, Hao Li, Yue Lu, Hongning Wang</i> | |
| Opinion Maximization in Social Networks..... | 387 |
| <i>Aristides Gionis, Evimaria Terzi, Panayiotis Tsaparas</i> | |
| Point-of-Interest Recommendation in Location Based Social Networks with Topic and Location Awareness..... | 396 |
| <i>Bin Liu, Hui Xiong</i> | |
| Community Detection with Prior Knowledge..... | 405 |
| <i>Charu Aggarwal, Jaideep Srivastava, Karthik Subbian, Philip Yu</i> | |

SESSION 10: CLASSIFICATION AND SPARSE METHODS

| | |
|---|-----|
| Regularization of Latent Variable Models to Obtain Sparsity..... | 414 |
| <i>Ramnath Balasubramanian, William W. Cohen</i> | |
| Sparse Max-Margin Multiclass and Multi-label Classifier Design for Fast Inference..... | 423 |
| <i>Tanuja Ganu, Sundararajan Sellamanickam, Shirish Shevade</i> | |
| An Empirical Study of the Suitability of Class Decomposition for Linear Models: When Does It Work Well?..... | 432 |
| <i>Francisco Ocegueda-Hernandez, Ricardo Vilalta</i> | |
| Reduced Set KPCA for Improving the Training and Execution Speed of Kernel Machines..... | 441 |
| <i>Alexander Gray, Hassan Kingravi, Patricio Vela</i> | |
| A New Perspective on Convex Relaxations of Sparse SVM..... | 450 |
| <i>Noam Goldberg, Sven Leyffer, Todd Munson</i> | |

POSTERS

| | |
|---|-----|
| Time-sensitive Classification of Behavioral Data..... | 458 |
| <i>Shin Ando, Einoshin Suzuki</i> | |
| An Examination of Practical Granger Causality Inference..... | 467 |
| <i>Yan Liu, Mohammad Taha Bahadori</i> | |
| Bregman Divergences and Triangle Inequality..... | 476 |
| <i>Sreangsu Acharyya, Arindam Banerjee, Daniel Boley</i> | |
| Joint Segmentation and Clustering in Text Corpuses..... | 485 |
| <i>Samuel Blasiak, Huzefa Rangwala, Sithu Sudarsan</i> | |
| Automatic Detection and Correction of Multi-class Classification Errors Using System Whole-part Relationships..... | 494 |
| <i>Zhengzhang Chen, Alok Choudhary, John Jenkins, Vipin Kumar, Anatoli Melechko, Jinfeng Rao, Nagiza Samatova, Fredrick Semazzi</i> | |
| Contextual Time Series Change Detection..... | 503 |
| <i>Shyam Boriah, Snigdhasu Chatterjee, Xi Chen, Vipin Kumar, Karsten Steinhaeuser</i> | |
| Very Fast Similarity Queries on Semi-Structured Data from the Web..... | 512 |
| <i>William Cohen, Bhavana Dalvi</i> | |
| Topic Models For Feature Selection in Document Clustering..... | 521 |
| <i>Anna Drummond, Chris Jermaine, Zografoula Vagena</i> | |
| A Nonparametric Mixture Model for Topic Modeling over Time..... | 530 |
| <i>Avinava Dubey, Ahmed Hefny, Sinead Williamson, Eric Xing</i> | |
| Discriminative Transfer Learning on Manifold..... | 539 |
| <i>Zheng Fang, Zhongfei Zhang</i> | |
| SemInf: A Burst-based Semantic Influence Model for Biomedical Topic Influence..... | 548 |
| <i>Dan He, Douglas S. Parker</i> | |
| Pinch Ratio Clustering from a Topologically Intrinsic Lexicographic Ordering..... | 560 |
| <i>Douglas Heisterkamp, Jesse Johnson</i> | |
| Retweeting: An Act of Viral Users, Susceptible Users, or Viral Topics?..... | 569 |
| <i>Ee-Peng Lim, Tuan-Anh Hoang</i> | |
| Time Series Classification under More Realistic Assumptions..... | 578 |
| <i>Yanping Chen, Bing Hu, Eamonn Keogh</i> | |
| Finding Affordable and Collaborative Teams from a Network of Experts..... | 587 |
| <i>Aijun An, Mehdi Kargar, Morteza Zihayat</i> | |

| | |
|---|-----|
| IBSM: Interval-Based Sequence Matching | 596 |
| <i>Vassilis Athitsos, Alexios Kotsifakos, Panagiotis Papapetrou</i> | |
| Modeling the Diffusion of Preferences on Social Networks | 605 |
| <i>San-Chuan Hung, Perng-Hwa Kung, Shou-De Lin, Jing-Kai Lou, Chin-Hua Tsai, Fu-Min Wang</i> | |
| Mining Labelled Tensors by Discovering both their Common and Discriminative Subspaces | 614 |
| <i>James Bailey, Jeffrey Chan, Ramamohanarao Kotagiri, Christopher Leckie, Wei Liu</i> | |
| Modeling Clinical Time Series Using Gaussian Process Sequences | 623 |
| <i>Milos Hauskrecht, Zitao Liu, Lei Wu</i> | |
| Integrity Verification of K-means Clustering Outsourced to Infrastructure as a Service (IaaS) Providers | 632 |
| <i>Ruilin Liu, Philippos Mordohai, Wendy Hui Wang, Hui Xiong</i> | |
| Selective Transfer Learning for Cross Domain Recommendation | 641 |
| <i>Zhongqi Lu, Weike Pan, Evan Wei Xiang, Qiang Yang, Lili Zhao, Erheng Zhong</i> | |
| Change Detection from Temporal Sequences of Class Labels: Application to Land Cover Change Mapping | 650 |
| <i>Shyam Boriah, Ankush Khandelwal, Vipin Kumar, Varun Mithal, Karsten Steinhäuser</i> | |
| Fractional Immunization in Networks | 659 |
| <i>Lada Adamic, Christos Faloutsos, Theodore Iwashyna, B. Aditya Prakash, Hanghang Tong</i> | |
| Fast Shapelets: A Scalable Algorithm for Discovering Time Series Shapelets | 668 |
| <i>Eamonn Keogh, Thanawin Rakthanmanon</i> | |
| MC-MinH: Metagenome Clustering using Minwise based Hashing | 677 |
| <i>Huzefa Rangwala, Zeehasham Rasheed</i> | |
| Shattering and Compressing Networks for Betweenness Centrality | 686 |
| <i>Umit V. Catalyurek, Kamer Kaya, Ahmet Erdem Sariyuce, Erik Saule</i> | |
| CoSelect: Feature Selection with Instance Selection for Social Media Data | 695 |
| <i>Huan Liu, Jiliang Tang</i> | |
| A Hierarchical Probabilistic Model for Low Sample Rate Home-Use Energy Disaggregation | 704 |
| <i>Arnold Boediardjo, Feng Chen, Haili Dong, Chang-Tien Lu, Bingsheng Wang</i> | |
| On the Detectability of Node Grouping in Networks | 713 |
| <i>Yuguo Chen, Jiawei Han, Ming Ji, Jialu Liu, Lu Su, Chi Wang, Hongning Wang</i> | |
| MODS: Multiple One-class Data Streams Learning from Homogeneous Data | 722 |
| <i>Zhifeng Hao, Bo Liu, Yanshan Xiao, Philip Yu</i> | |
| Robust Textual Data Streams Mining Based on Continuous Transfer Learning | 731 |
| <i>Longbing Cao, Zhifeng Hao, Bo Liu, Yanshan Xiao, Philip Yu</i> | |
| Graphical Modeling of Macro Behavioral Targeting in Social Networks | 740 |
| <i>Ankit Agrawal, Zhengzhang Chen, Yu Cheng, Alok Choudhary, Haotian Liu, Md. Mostofa Ali Patwary, Yusheng Xie, Kumpeng Zhang</i> | |
| Learning Topics in Short Texts by Non-negative Matrix Factorization on Term Correlation Matrix | 749 |
| <i>Xueqi Cheng, Jiafeng Guo, Shenghua Liu, Yanfeng Wang, Xiaohui Yan</i> | |
| Set Coverage Problems in a One-Pass Data Stream | 758 |
| <i>Huiwen Yu, Dayu Yuan</i> | |
| Sentiment Topic Model with Decomposed Prior | 767 |
| <i>Zheng Chen, Chengtao Li, Jian-Tao Sun, Jianwen Zhang</i> | |
| Topic-Level Expert Modeling in Community Question Answering | 776 |
| <i>Naiwen Bian, Chungping Li, Mengya Li, Tong Zhao</i> | |
| Butterfly Mixing: Accelerating Incremental-Update Algorithms on Clusters | 785 |
| <i>John Canny, Huasha Zhao</i> | |
| It Is Not Just What We Say, But How We Say Them: LDA-based Behavior-Topic Model | 794 |
| <i>Jing Jiang, Minghui Qiu, Feida Zhu</i> | |
| Feature Selection by Joint Graph Sparse Coding | 803 |
| <i>Wei Ding, Xindong Wu, Shichao Zhang, Xiaofeng Zhu</i> | |
| Author Index | |