

Sixteenth SIAM International Conference on Data Mining 2016 (SDM16)

Miami, Florida
5 - 7 May 2016

Editors:

**Sanjay Chawla Venkatasubramanian
Wagner Meira**

ISBN: 978-1-5108-2811-7

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© (2016) by SIAM: Society for Industrial and Applied Mathematics
All rights reserved.

Printed by Curran Associates, Inc. (2016)

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

New Methods and Applications

Clustering in the Face of Fast Changing Streams.....	1
Liudmila Ulanova; Nurjahan Begum; Mohammad Shokoohi-Yekta; Eamonn Keogh	
Kernelized Sparse Self-Representation for Clustering and Recommendation	10
Xiao Bian; Feng Li; Xia Ning	
Multi-Domain Manifold Learning for Drug-Target Interaction Prediction	18
Ruichu Cai; Zhenjie Zhang; Srinivasan Parthasarathy; Anthony K. H. Tung; Zhifeng Hao; Wen Zhang	
Finding Surprisingly Frequent Patterns of Variable Lengths in Sequence Data	27
Reza Sadoddin; Joerg Sander; Davood Rafiei	
Identifying Connectivity Patterns for Brain Diseases via Multi-side-view Guided Deep Architectures.....	36
Jingyuan Zhang; Bokai Cao; Sihong Xie; Chun-ta Lu; Philip Yu; Ann Ragin	
Regularized Weighted Linear Regression for High-dimensional Censored Data.....	45
Yan Li; Bhanukiran Vinzamuri; Chandan Reddy	

Network Analysis

Node Classification in Signed Social Networks	54
Jiliang Tang; Charu Aggarwal; Huan Liu	
Uncovering Multiple Diffusion Networks Using the First-Hand Sharing Pattern.....	63
Pei-Lun Liao; Chung-Kuang Chou; Ming-Syan Chen	
Integrating Community and Role Detection in Information Networks	72
Ting Chen; Yizhou Sun; Lu-an Tang; Zhengzhang Chen; Haifeng Chen; Guofei Jiang	
Exploiting Emotional Information for Trust/Distrust Prediction	81
Ghazaleh Beigi; Jiliang Tang; Suhan Wang; Huan Liu	
Online Prediction of User Actions through an Ensemble Vote from Vector Representation and Frequency Analysis Models	90
Changsung Moon; Dakota Medd; Paul Jones; Nagiza Samatova; Steve Harenberg; William Oxbury	
FairPlay: Fraud and Malware Detection in Google Play	99
Mahmudur Rahman; Mizanur Rahman; Bogdan Carbunar; Polo Chau	

Recommendation Systems and Factorization

Synergies that Matter: Efficient Interaction Selection via Sparse Factorization Machine.....	108
Jianpeng Xu; Kaixiang Lin; Pang-Ning Tan; Jiayu Zhou	
A Spatial-Temporal Probabilistic Matrix Factorization Model for Point-of-Interest Recommendation.....	117
Huayu Li; Yong Ge; Richang Hong; Zhiang Wu	
Top-N Recommendation with Novel Rank Approximation.....	126
Zhao Kang; qiang Cheng	
Vocal Competence Based Karaoke Recommendation: A Maximum-Margin Joint Model.....	135
Chu Guan; Yanjie Fu; Xinjiang Lu; Hui Xiong; Enhong Chen; Yingling Liu	
A Confidence-Based Approach for Balancing Fairness and Accuracy.....	144
Benjamin Fish; Jeremy Kun; Adam Lelkes	
Differentially Private Statistical Significance Testing on Paired Sample Data	153
Chris Clifton; Christine Task	

Outlier Detection

A General Framework to Increase the Robustness of Model-Based Change Point Detection Algorithms to Outliers and Noise	162
Xi Chen; Yuanshun Yao; Sichao Shi; Snigdhanu Chatterjee; Vipin Kumar; James Faghmous	
LODES: Local Density Meets Spectral Outlier Detection	171
Saket Sathe; Charu Aggarwal	
Routine Mining Based Anomaly Detection in Mobile Phone Data.....	180
Tian Qin; Guojie Song; Sizhen Du	
A Scalable Approach for Outlier Detection in Edge Streams Using Sketch-based Approximations.....	189
Stephen Ranshous; Steve Harenberg; Kshitij Sharma; Nagiza Samatova; Kshitij Sharma	
R1STM: One-class Support Tensor Machine with Randomised Kernel.....	198
Sarah Erfani; Mahsa Baktashmotlagh; Sutharshan Rajasegar; Vinh Nguyen; Christopher Leckie; James Bailey; Kotagiri Ramamohanar	
Scalable Anomaly Ranking of Attributed Neighborhoods.....	207
Bryan Perozzi; Leman Akoglu	

Advanced Classification

Linear and Kernel Classification: When to Use Which?	216
Hsin Yuan Huang; Chih-Jen Lin	
Pattern Aided Classification	225
Guozhu Dong; Vahid Taslimitehrani	
TerseSVM : A Scalable Approach for Learning Compact Models in Large-scale Classification	234
Rohit Babbar; Krikamol Maundet; bernhard Schölkopf	
Discriminative Training of Structured Dictionaries via Block Orthogonal Matching Pursuit.....	243
Wenling Shang; Honglak Lee; Anna Gilbert; Kihyuk Sohn; Anna Gilbert	
A Unified View of Localized Kernel Learning	252
John Moeller; Sarathkrishna Swaminathan; Suresh Venkatasubramiam	
Binary Classifier Calibration Using an Ensemble of Linear Trend Estimation	261
Mahdi Pakdaman Naeini; Gregory Cooper	

Clustering and Embedding

Power Simultaneous Spectral Data Embedding and Clustering	270
Kais Allab; Lazhar Labiod; Mohamed Nadif	
Process Trace Clustering: A Heterogeneous Information Network Approach.....	279
Phuong Nguyen; Aleksander Slominski; Vinod Muthusamy; Vatche Ishakian; Klara Nahrstedt	
Lagrangian Constrained Clustering.....	288
Mohadeseh Ganji; James Bailey; Peter Stuckey	
Fast Multiplier Methods to Optimize Non-exhaustive, Overlapping Clustering	297
Yangyang Hou; Joyce Whang; David Gleich; Inderjit Dhillon	
Stochastic Co-clustering for Document-Term Data.....	306
Aghiles SALAH; Nicoleta Rogovschi; Mohamed Nadif	
On Finding the Maximum Edge Biclique in a Bipartite Graph: a Subspace Clustering Approach	315
Eran Shaham; Honghai Yu; Xiao-Li Li	

Scaling Clustering Algorithms

Geometric Methods to Accelerate K-Means Algorithms	324
Petr Ryšavý; Greg Hamerly	

Pivot-based k-means Algorithm for Numerous-class Data Sets	333
Takashi Hattori; Kazuo Aoyama; Kazumi Saito; Tetsuo Ikeda; Eri Kobayashi	
k-Means for Streaming and Distributed Big Sparse Data	342
Artem Barger; Danny Feldman	
Haliteds: Fast and Scalable Subspace Clustering for Multidimensional Data Streams	351
Afonso E. da Silva; lucas Lancellotti Sanches; Antonio Canabrava Fraideinberze; Robson Cordeiro	
Online Clustering of Multivariate Time-series	360
Masud Moshtaghi; Christopher Leckie; James Bezdek	
Learning A Task-Specific Deep Architecture For Clustering	369
Zhangyang Wang; Shiyu Chang; Jiayu Zhou; Meng Wang; Thomas Huang	
Feature Selection	
Kernelized Matrix Factorization for Collaborative Filtering.....	378
Xinyue Liu; Charu Aggarwal; Yu-Feng Li; Xiangnan Kong; Xinyuan Sun; Saket Sathe	
Robust Unsupervised Feature Selection on Networked Data	387
Jundong Li; Xia Hu; Harsh Dani; Liang Wu; Huan Liu	
Euclidean Co-Embedding of Ordinal Data for Multi-Type Visualization.....	396
Dung Le; Hady Lauw	
K-Nearest Neighbor Search and Outlier Detection via Minimax Distances.....	405
Morteza Haghir Chehreghani	
Nonlinear Joint Unsupervised Feature Selection	414
Xiaokai Wei; Bokai Cao; Philip Yu	
A Framework to Adjust Dependency Measure Estimates for Chance	423
Simone Romano; Vinh Nguyen; James Bailey; Karin Verspoor	
Applications	
Risk Prediction with Electronic Health Records: A Deep Learning Approach.....	432
Yu Cheng; Fei Wang; Ping Zhang; Jianying Hu	
Predicting the Popularity of News Articles.....	441
Yaser Keneshloo; Shuguang Wang; Sam Han; Naren Ramakrishnan	

Uncovering Latent Behaviors in Ant Colonies	450
Mohamed Kafsi; Patrick Thiran; Matthias Grossglauser; Danielle Mersch; Raphael Braunschweig; Laurent Keller	
The Impact of Community Safety on House Ranking	459
Zijun Yao; Yanjie Fu; Bin Liu; Hui Xiong	
iPath: Forecasting the Pathway to Impact	468
Liangyue Li; Hanghang Tong; Jie Tang; Wei Fan	
Cost-Sensitive Batch Mode Active Learning: Designing Astronomical Observation by Optimizing Telescope Time and Telescope Choice.....	477
Xide Xia; Pavlos Protopapas; Finale Doshi-Velez	
 Graphs, Networks, Communities	
A Fast Kernel for Attributed Graphs	486
Yu Su; Fangqiu Han; Richard Harang; Xifeng Yan	
BIRDNEST: Bayesian Inference for Ratings-Fraud Detection	495
Bryan Hooi; Neil Shah; Alex Beutel; Stephan Günnemann; Leman Akoglu; Mohit Kumar; Disha Makhija; Christos Faloutsos	
Unstable Communities in Network Ensembles	504
Md. Ahsanur Rahman; Steve T. K. Jan; Hyunju Kim; B. Aditya Prakash; T. M. Murali	
CAML P: Confidence-Aware Modulated Label Propagation	513
Yuto Yamaguchi; Christos Faloutsos; Hiroyuki Kitagawa	
Query-Driven Maximum Quasi-Clique Search	522
Pei Li; Laks Lakshmanan	
Distributed Representations of Expertise	531
Fangqiu Han; shulong Tan; Huan Sun; Mudhakar Srivatsa; Deng Cai; Xifeng Yan	
 Temporal, Stream and Pattern Mining	
Temporal Kernel Descriptors for Learning with Time-sensitive Patterns	540
Doyen Sahoo; Abhishek Sharma; Steven Hoi; Peilin Zhao	
Modelling Recurrent Events for Improving Online Change Detection	549
Alexandr Maslov; Mykola Pechenizkiy; Tommi Karkkainen; Indre Zliobaite	
MACFP: Maximal Approximate Consecutive Frequent Pattern Mining under Edit Distance.....	558
Jingbo Shang; Jian Peng; Jiawei Han	

DPClass: An Effective but Concise Discriminative Patterns-Based Classification Framework	567
Jingbo Shang; Wenzhu Tong; Jian Peng; Jiawei Han	
Fast Lossless Frequent Itemset Mining in Data Streams using Crucial Patterns	576
Ariyam Das; Carlo Zaniolo	
Flexibly Mining Better Subgroups	585
Hoang-Vu Nguyen; Jilles Vreeken	
 Inference with Optimization	
Deterministic Column Sampling for Low-Rank Matrix Approximation: Nystrom vs. Incomplete Cholesky Decomposition	594
Tom Goldstein; Raajen Patel; Eva Dyer; Richard Baraniuk; Azalia Mirhoseini	
A Polynomial Expansion Line Search for Large-Scale Unconstrained Minimization of Smooth L2-Regularized Loss Functions, With Implementation in Apache Spark	603
Michael Hynes; Hans De Sterck	
Structured Regression on Multilayer Networks.....	612
Athanasia Polychronopoulou; Zoran Obradovic	
RelSim: Relation Similarity Search in Schema-Rich Heterogeneous Information Networks	621
Chenguang Wang; Yizhou Sun; Yanglei Song; Jiawei Han; Yangqiu Song; lidan Wang; Ming Zhang	
Collective Opinion Spam Detection using Active Inference.....	630
Shebuti Rayana; Leman Akoglu	
Discovery of Precursors to Adverse Events using Time Series Data.....	639
Vijay Manikandan Janakiraman; Bryan Matthews; Nikunj Oza	
 Structured Learning	
Effective Crowd Expertise Modeling via Cross Domain Sparsity and Uncertainty Reduction	648
Sihong Xie; Qingbo Hu; Weixiang Shao; Jingyuan Zhang; Jing Gao; Wei Fan; Philip Yu	
GSpartan: a Geospatio-Temporal Multi-task Learning Framework for Multi-location Prediction	657
Jianpeng Xu; Pang-Ning Tan; Lifeng Luo; Jiayu Zhou	

Learning Correlative and Personalized Structure for Online Multi-Task Classification	666
Peng Yang; Peilin Zhao	
Online Sparse Passive Aggressive Learning with Kernels	675
Jing Lu; Peilin Zhao; Steven Hoi	
ADMM for Training Sparse Structural SVMs with Augmented Regularizers	684
Balamurugan Palaniappan; Anusha Posinasetty; Shirish Shevade	
Structural Orthogonal Procrustes Regression for Face Recognition with Pose Variations and Misalignment.....	693
Ying Tai; Jian Yang; Fanlong Zhang; Yigong Zhang; Lei Luo; Jianjun Qian	
Probabilistic Inference	
Capricorn: An Algorithm for Subtropical Matrix Factorization	702
Sanjar Karaev; Pauli Miettinen	
Automatic Unsupervised Tensor Mining with Quality Assessment	711
Evangelos Papalexakis	
Rank Selection for Non-negative Matrix Factorization Using Normalized Maximum Likelihood Coding.....	720
Yu Ito; Shin-ichi Oeda; Kenji Yamanishi	
Sparse Hybrid Variational-Gibbs algorithm for Latent Dirichlet Allocation.....	729
Ximing Li; Jihong Ouyang; Xiaotang Zhou	
Scaling Lifted Probabilistic Inference and Learning Via Graph Databases	738
Mayukh Das; Yuqing Wu; Tushar Khot; Kristian Kersting; Sriraam Natarajan	
Estimating Posterior Ratio for Classification: Transfer Learning from Probabilistic Perspective	747
Song Liu; Kenji Fukumizu	
High-dimensional Analysis	
Constrained Group Testing to Predict Binding Response of Candidate Compounds.....	756
Paul Quint; Stephen Scott; N Vinodchandran; Brad Worley	
Regularized Parametric Regression for High-dimensional Survival Analysis.....	765
Yan Li; Kevin Xu; Chandan Reddy	
Copula-HDP-HMM: Non-parametric Modeling of Temporal Multivariate Data for I/O Efficient Bulk Cache Preloading.....	774
Lavanya Sita Tekumalla; Chiranjib Bhattacharyya	

On Skewed Multi-dimensional Distributions: the FusionRP Model, Algorithms, and Discoveries	783
Venkata Krishna Pillutla; Zhanpeng Fang; Pravallika Devineni; Jie Tang; Christos Faloutsos; Danai Koutra	
Universal Dependency Analysis.....	792
Hoang-Vu Nguyen; Panagiotis Mandros; Jilles Vreeken	
High Dimensional Structured Estimation with Noisy Designs	801
Amir Asiaee T.; Soumyadeep Chatterjee; Arindam Banerjee	
Spatio-temporal Mining	
Learning Linear Dynamical Systems from Multivariate Time Series: A Matrix Factorization Based Framework	810
Zitao Liu; Milos Hauskrecht	
Spatio-Temporal Tensor Analysis for Whole-Brain fMRI Classification	819
Guixiang Ma; Lifang He; Chun-ta Lu; Philip Yu; Linlin Shen; Ann Ragin	
Linear-time Detection of Non-linear Changes in Massively High Dimensional Time Series	828
Hoang-Vu Nguyen; Jilles Vreeken	
Speeding Up All-Pairwise Dynamic Time Warping Matrix Calculation.....	837
Diego Furtado Silva; Gustavo Batista	
Joint Learning of Representation and Structure for Sparse Regression on Graphs.....	846
Chao Han; Shanshan Zhang; Mohamed Ghalwash; Slobodan Vucetic; Zoran Obradovic	
Infusing Geo-Recency Mixture Models for Effective Location Prediction in LBSN	855
Roland Assam; Subramanyam Sathyanarayana; Thomas Seidl	