

Seventeenth SIAM International Conference on Data Mining 2017 (SDM17)

Houston, Texas, USA
27 - 29 April 2017

Editors:

**Nitesh Chawla
Wei Wang**

ISBN: 978-1-5108-4442-1

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© (2017) by SIAM: Society for Industrial and Applied Mathematics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

Session 1: Classification I

Using a Random Forest to Inspire a Neural Network and Improving on It	1
Suhang Wang, Charu Aggarwal, Huan Liu	
Pruning Decision Trees via Max-Heap Projection	10
Zhi Nie, Binbin Lin, Shuai Huang, Naren Ramakrishnan, Wei Fan, Jieping Ye	
Margin Distribution Logistic Machine	19
Yi Ding, Sheng-Jun Huang, Chen Zu, Daoqiang Zhang	
Active Learning of Classification Models with Likert-Scale Feedback	28
Yanbing Xue, Milos Hauskrecht	
Risk Clearance with Guaranteed Precision	36
Ryan McBride, Ke Wang, Viswanadh Nekkanti, Wenyuan Li	

Session 2: Bioinformatics

Computational Drug Discovery with Dyadic Positive-Unlabeled Learning	45
Yashu Liu, Shuang Qui, Ping Zhang, Pinghua Gong, Fei Wang, Guoliang Xue, Jieping Ye	
Multi-Region Neural Representation: A Novel Model for Decoding Visual Stimuli in Human Brains	54
Muhammad Yousefnezhad, Daoqiang Zhang	
ALPINE: Progressive Itemset Mining with Definite Guarantees	63
Qiong Hu, Tomasz Imielinski	
Polyadic Regression and Its Application to Chemogenomics	72
Ioakeim Perros, Fei Wang, Ping Zhang, Peter Walker, Richard Vuduc, Jyotishman Pathak, Jimeng Sun	
Active Learning of Functional Networks from Spike Trains	81
Honglei Liu, Bian Wu	

Session 3: Anomaly and Outlier Detection

Outlier Detection with Autoencoder Ensembles	90
Jinghui Chen, Saket Sathe, Charu Aggarwal, Deepak Turaga	
Gleaning Wisdom from the Past: Early Detection of Emerging Rumors in Social Media	99
Liang Wu, Jundong Li, Xia Hu, Huan Liu	

VolTime: Unsupervised Anomaly Detection on Users' Online Activity Volume	108
Daniel Y.T. Chino, Alceu F. Costa, Agma J.M. Traina, Christos Faloutsos	
Detecting Malicious Behavior in Computer Networks via Cost-Sensitive and Connectivity Constrained Classification	117
Houping Xiao, Jin Gao, Long Vu, Deepak S. Turaga	
Efficiently Discovering Unexpected Pattern-Co-Occurrences	126
Roel Bertens, Jilles Vreeken, Arno Siebes	

Session 4: Classification II

Time-Aware Subscription Prediction Model for User Acquisition in Digital News Media	135
Heidar Davoudi, Morteza Zihayat, Aijun An	
The Power of Certainty: A Dirichlet-Multinomial Model for Belief Propagation	144
Dhivya Eswaran, Stephan Günnemann, Christos Faloutsos	
From Theory to Practice: Efficient Active Cost-sensitive Classification with Expected Error Reduction	153
Yexun Zhang, Yanfeng Wang, Wenbin Cai, Siyuan Zhou, Ya Zhang	
Generalized Inverse Classification	162
Michael T. Lash, Qihang Lin, Nick Street, Jennifer G. Robinson, Jeffrey Ohlmann	
Predict Land Covers with Transition Modeling and Incremental Learning	171
Xiaowei Jia, Ankush Khandelwal, Guruprasad Nayak, James Gerber, Kimberly Carlson, Paul West, Vipin Kumar	

Session 5: Health Informatics

Unified and Contrasting Graphical Lasso for Brain Network Discovery	180
Xinyue Liu, Xiangnan Kong, Ann B. Ragin	
t-BNE: Tensor-based Brain Network Embedding	189
Bokai Cao, Lifang He, Xiaokai Wei, Mengqi Xing, Philip S. Yu, Heide Klumpp, Alex D. Leow	
An RNN Architecture with Dynamic Temporal Matching for Personalized Predictions of Parkinson's Disease	198
Chao Che, Cao Xiao, Jian Liang, Bo Jin, Jiayu Zhou, Fei Wang	
Clustering with Domain-Specific Usefulness Scores	207
Yale Chang, Junxiang Chen, Michael H. Cho, Peter J. Castaldi, Edwin K. Silverman, Jennifer G. Dy	
BRAINZOOM: High Resolution Reconstruction from Multi-modal Brain Signals	216
Xiao Fu, Kejun Huang, Otilia Stretcu, Hyun Ah Song, Evangelos Papalexakis, Partha Talukdar, Tom Mitchell, Nicholas Sidiropoulos, Christos Faloutsos, Barnabas Poczos	

Session 6: Clustering and Recommender Systems

Uncovering Group Level Insights with Accordant Clustering	228
Amit Dhurandhar, Margareta Ackerman, Xiang Wang	
A Method to Accelerate Human in the Loop Clustering	237
Anni Coden, Marina Danilevsky, Daniel Gruhl, Linda Kato, Meena Nagarajan	
Model-based von Mises-Fisher Co-clustering with a Conscience	246
Aghiles Salah, Mohamed Nadif	
Targeted Matrix Completion	255
Natali Ruchansky, Mark Crovella, Evimaria Terzi	
Active Positive-Definite Matrix Completion	264
Charalampos Mavroforakis, Dóra Erdős, Mark Crovella, Evimaria Terzi	

Session 7: Scalable Data Mining

Multi-core K-means	273
Christian Böhm, Martin Perdacher, Claudia Plant	
Indexing and Classifying Gigabytes of Time Series under Time Warping	282
Chang Wei Tan, Geoffrey I. Webb, François Petitjean	
A Sparse Nonlinear Classifier Design using AUC Optimization	291
Vishal Kakkar, Shirish Shevade, S. Sundararajan, Dinesh Garg	
A Dual-Tree Algorithm for Fast k-means Clustering with Large k	300
Ryan R. Curtin	
Specious Rules: An Efficient and Effective Unifying Method for Removing Misleading and Uninformative Patterns in Association Rule Mining	309
Wilhelmiina Hämäläinen, Geoffrey I. Webb	

Session 8: Mining Graphs I

MeiKe: Influence-based Communities in Networks	318
Yao Zhang, Bijaya Adhikari, Steve T.K. Jan, B. Aditya Prakash	
Signed Network Embedding in Social Media	327
Suhang Wang, Jiliang Tang, Charu Aggarwal, Yi Chang, Huan Liu	
Finding Low-Tension Communities	336
Esther Galbrun, Behzad Golshan, Aristides Gionis, Evimaria Terzi	
Sensitivity of Community Structure to Network Uncertainty	345
Marc Mitri, Fragkiskos D. Malliaros, Michalis Vazirgiannis	

Subnetwork Mining with Spatial and Temporal Smoothness	354
Xuan-Hong Dang, Hongyuan You, Ambuj K. Singh, Scott Grafton	

Session 9: Recommender Systems

Selection of Negative Samples for One-class Matrix Factorization	363
Hsiang-Fu Yu, Mikhail Bilenko, Chih-Jen Lin	

HBGG: A Hierarchical Bayesian Geographical Model for Group Recommendation	372
Ziyu Lu, Hui Li, Nikos Mamoulis, David W. Cheung	

Collaborative User Network Embedding for Social Recommender Systems	381
Chuxu Zhang, Lu Yu, Yan Wang, Chirag Shah, Xiangliang Zhang	

Redundancies in Data and their Effect on the Evaluation of Recommendation Systems: A Case Study on the Amazon Reviews Datasets	390
Daniel Basaran, Eirini Ntoutsis, Arthur Zimek	

Price Recommendation on Vacation Rental Websites	399
Yang Li, Suhang Wang, Tao Yang, Quan Pan, Jiliang Tang	

Session 10: Social Network Analysis

Ranking in Heterogeneous Networks with Geo-Location Information	408
Abhinav Mishra, Leman Akoglu	

Condensing Temporal Networks using Propagation	417
Bijaya Adhikari, Yao Zhang, Aditya Bharadwaj, B. Aditya Prakash	

Community-aware Network Sparsification	426
Aristides Gionis, Polina Rozenshtein, Nikolaj Tatti, Evimaria Terzi	

Graph-based Semi-Supervised Learning for Relational Networks	435
Leto Peel	

Toward Personalized Relational Learning	444
Jundong Li, Liang Wu, Osmar R. Zaiane, Huan Liu	

Session 11: Text and Web Mining

Biclustering: An Application of Dual Topic Models	453
Daniel Rugeles, Kaiqi Zhao, Cong Gao, Manoranjan Dash, Shonali Krishnaswamy	

A Graduated Non-Convexity Relaxation for Large Scale Seriation	462
Xenophon Evangelopoulos, Austin J. Brockmeier, Tingting Mu, John Y. Goulermas	

User-guided Cross-domain Sentiment Classification	471
Arun Reddy Nelakurthi, Hanghang Tong, Ross Maciejewski, Nadya Bliss, Jingrui He	

Exploring Latent Semantic Factors to Find Useful Product Reviews 480
Subhabrata Mukherjee, Kashyap Popat, Gerhard Weikum

Outlier Detection for Text Data 489
Ramakrishnan Kannan, Hyenkyun Woo, Charu C. Aggarwal, Haesun Park

Session 12: Feature, Extraction, Selection, and Dimension Reduction

ROFLMAO: Robust Oblique Forests with Linear Matrix Operations 498
Tyler M. Tomita, Mauro Maggioni, Joshua T. Vogelstein

Exploiting Hierarchical Structures for Unsupervised Feature Selection 507
Suhang Wang, Yilin Wang, Jiliang Tang, Charu Aggarwal, Suhas Ranganath, Huan Liu

Embedded Supervised Feature Selection for Multi-class Data 516
Lin Chen, Jiliang Tang, Baoxin Li

Correlation by Compression 525
Kailash Budhathoki, Jilles Vreeken

A Deflation Method for Structured Probabilistic PCA 534
Rajiv Khanna, Joydeep Ghosh, Russell Poldrack, Oluwasanmi Koyejo

Session 13: Novel Applications

Identifying Deep Contrasting Networks from Time Series Data: Application to Brain Network Analysis 543
John Boaz Lee, Xiangnan Kong, Yihan Bao, Constance Moore

Cumulative Knowledge-based Regression Models for Next-term Grade Prediction 552
Sara Morsy, George Karypis

BreachRadar: Automatic Detection of Points-of-Compromise 561
Miguel Araujo, Miguel Almeida, Jaime Ferreira, Luis Silva, Pedro Bizarro

HiDDen: Hierarchical Dense Subgraph Detection with Application to Financial Fraud Detection 570
Si Zhang, Dawei Zhou, Mehmet Yigit Yildirim, Scott Alcorn, Jingrui He, Hasan Davulcu, Hanghang Tong

MultiC²: An Optimization Framework for Learning from Task and Worker Dual Heterogeneity 579
Yao Zhou, Lei Ying, Jingrui He

Uplift Modeling with Multiple Treatments and General Response Types 588
Yan Zhao, Xiao Fang, David Simchi-Levi

Session 14: Mining Graphs II

FACETS: Adaptive Local Exploration of Large Graphs	597
Robert Pienta, Minsuk Kahng, Zhiyuan Lin, Jilles Vreeken, Partha Talukdar, James Abello, Ganesh Parameswaran, Duen Horng Chau	
Absenteeism Detection in Social Media	606
Fang Jin, Feng Chen, Rupinder Paul Khandpur, Chang-Tien Lu, Naren Ramakrishnan	
Multimodal Network Alignment	615
Huda Nassar, David F. Gleich	
Near-Optimal and Practical Algorithms for Graph Scan Statistics	624
Jose Cadena, Feng Chen, Anil Vullikanti	
Accelerated Attributed Network Embedding	633
Xiao Huang, Jundong Li, Xia Hu	
Meta-Path Graphical Lasso for Learning Heterogeneous Connectivities	642
Yao Zhang, Yun Xiong, Xinyue Liu, Xiangnan Kong, Yangyong Zhu	

Session 15: Novel Learning Methods

Statistical Learning Theory Approach for Data Classification with I-Diversity	651
Koray Mancuhan, Chris Clifton	
Private and Right-Protected Big Data Publication: An Analysis	660
Reinhard Heckel, Michail Vlachos	
Differentially Private Rank Aggregation	669
Michael Hay, Liudmila Elagina, Gerome Miklau	
Hash-Based Feature Learning for Incomplete Continuous-Valued Data	678
Shuai Yuan, Pang-Ning Tan, Kendra S. Cheruvilil, C. Emi Fergus, Nicholas K. Skaff, Patricia A. Soranno	
Multi-Task Multiple Kernel Relationship Learning	687
Keerthiram Murugesan, Jaime Carbonell	
Multivariate Confidence Intervals	696
Jussi Korpela, Emilia Oikarinen, Kai Puolamäki, Antti Ukkonen	

Session 16: Optimization Methods

A Fast Trust-Region Newton Method for Softmax Logistic Regression	705
Nayyar A. Zaidi, Geoffrey I. Webb	
Learning from Multi-Modality Multi-Resolution Data: An Optimization Approach	714
Yada Zhu, Jianbo Li, Jingrui He	

Sparse Graphical Modeling via Stochastic Complexity	723
Kohei Miyaguchi, Shin Matsushima, Kenji Yamanishi	
Limited-memory Common-directions Method for Distributed Optimization and its Application on Empirical Risk Minimization	732
Ching-pei Lee, Po-Wei Wang, Weizhu Chen, Chih-Jen Lin	
Automatic Frankensteining: Creating Complex Ensembles Autonomously	741
Martin Wistuba, Nicolas Schilling, Lars Schmidt-Thieme	

Session 17: Data Stream Mining

Error Metrics for Learning Reliable Manifolds from Streaming Data	750
Frank Schoeneman, Suchismit Mahapatra, Varun Chandola, Nils Napp, Jaroslaw Zola	
CSTG: An Effective Framework for Cost-sensitive Sparse Online Learning	759
Zhong Chen, Zhide Fang, Wei Fan, Andrea Edwards, Kun Zhang	
Concept Drift Detection with Hierarchical Hypothesis Testing	768
Shujian Yu, Zubin Abraham	
Deep Learning: A Generic Approach for Extreme Condition Traffic Forecasting	777
Rose Yu, Yaguang Li, Cyrus Shahabi, Ugur Demiryurek, Yan Liu	
H-FUSE: Efficient Fusion of Aggregated Historical Data	786
Zongge Liu, Hyun Ah Song, Vladimir Zadorozhny, Christos Faloutsos, Nicholas Sidiropoulos	

Session 18: Spatial-Temporal Mining

Efficiently Summarising Event Sequences with Rich Interleaving Patterns	795
Apratim Bhattacharyya, Jilles Vreeken	
Discovery of Causal Time Intervals	804
Zhenhui Li, Guanjie Zheng, Amal Agarwal, Lingzhou Xue, Thomas Lauvaux	
Robust Map Matching for Heterogeneous Data via Dominance Decompositions	813
Martin P. Seybold	
Uncovering the Spatiotemporal Patterns of Collective Social Activity	822
Martin Jankowiak, Manuel Gomez-Rodriguez	
Discovering Bursts Revisited: Guaranteed Optimization of the Model Parameters	831
Nikolaj Tatti	