

2017 IEEE International Conference on Data Mining Workshops (ICDMW 2017)

**New Orleans, Louisiana, USA
18 - 21 November 2017**

Pages 1-596



**IEEE Catalog Number: CFP1756B-POD
ISBN: 978-1-5386-3801-9**

**Copyright © 2017 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP1756B-POD
ISBN (Print-On-Demand):	978-1-5386-3801-9
ISBN (Online):	978-1-5386-3800-2
ISSN:	2375-9232

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
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2017 IEEE International Conference on Data Mining Workshops

ICDMW 2017

Table of Contents

Message from the Conference Chairs.....	xvii
Message from the Workshop Co-Chairs.....	xix
Organizing Committee.....	xxi

DSBDA - Data Science and Big Data Analytics

Intent-Aware Contextual Recommendation System	1
<i>Biswarup Bhattacharya, Iftikhar Burhanuddin, Abhilasha Sancheti, and Kushal Satya</i>	
A Big Data Analysis Framework Using Apache Spark and Deep Learning	9
<i>Anand Gupta, Hardeo Kumar Thakur, Ritvik Shrivastava, Pulkit Kumar, and Sreyashi Nag</i>	
Analyzing Dynamical Activities of Co-occurrence Patterns for Cooking Ingredients	17
<i>Yuuki Kikuchi, Masahito Kumano, and Masahiro Kimura</i>	
Distance and Density Clustering for Time Series Data	25
<i>Ruizhe Ma and Rafal Angryk</i>	
Accelerated Hierarchical Density Based Clustering	33
<i>Leland McInnes and John Healy</i>	
Hybrid.poly: An Interactive Large-Scale In-memory Analytical Polystore	43
<i>Maksim Podkorytov, Dylan Soderman, and Michael Gubanov</i>	
Action Rules for Sentiment Analysis on Twitter Data Using Spark	51
<i>Jaishree Ranganathan, Allen S. Irudayaraj, and Angelina A. Tzacheva</i>	
Automated Storytelling Evaluation and Story Chain Generation	61
<i>J.T. Rigsby and Daniel Barbará</i>	
Dealing with Class Imbalance the Scalable Way: Evaluation of Various Techniques Based on Classification Grade and Computational Complexity	69
<i>Bernhard Schlegel and Bernhard Sick</i>	
Failure Prediction with Adaptive Multi-scale Sampling and Activation Pattern Regularization	79
<i>Yujin Tang, Shinya Wada, and Kiyohito Yoshihara</i>	
Discovery of Action Rules at Lowest Cost in Spark	87
<i>Angelina A. Tzacheva, Arunkumar Bagavathi, and Lavanya Ayila</i>	

Estimating Treatment Effects of a Residential Demand Response Program Using Non-experimental Data	95
<i>Datong Paul Zhou, Maximilian Balandat, and Claire Jennifer Tomlin</i>	

Deep and Confident Prediction for Time Series at Uber	103
<i>Lingxue Zhu and Nikolay Laptev</i>	

DaMNet - Data Mining in Networks

Distributed Representations of Subgraphs	111
<i>Bijaya Adhikari, Yao Zhang, Naren Ramakrishnan, and B. Aditya Prakash</i>	

Network Model Selection for Task-Focused Attributed Network Inference	118
<i>Ivan Brugere, Chris Kanich, and Tanya Y. Berger-Wolf</i>	

Mining Patterns of Cyberbullying on Twitter	126
<i>Charalampos Chelmis, Daphney-Stavroula Zois, and Mengfan Yao</i>	

Principled Multilayer Network Embedding	134
<i>Weiyi Liu, Pin-yu Chen, Sailung Yeung, Toyotaro Suzumura, and Lingli Chen</i>	

HUBify: Efficient Estimation of Central Entities Across Multiplex Layer Compositions	142
<i>Abhishek Santra, Sanjukta Bhowmick, and Sharma Chakravarthy</i>	

Predicting Users' Search Behavior Using Stochastic Multi-mode Network Models	150
<i>Shohei Umehara and Koji Eguchi</i>	

Constructing a Hierarchical User Interest Structure Based on User Profiles	156
<i>Chao Zhao, Min Zhao, and Yi Guan</i>	

Exploring Uncertainty Methods for Centrality Analysis in Social Networks	163
<i>Xianglin Zuo, Bo Yang, and Wanli Zuo</i>	

DSHCM - Data Science for Human Capital Management

Analysis of Organizational Structure Through Cluster Validation Techniques: Evaluation of Email Communications at an Organizational Level	170
<i>Veselka Boeva, Lars Lundberg, Sai M. Harsha Kota, and Lars Sköld</i>	

Apollo: Near-Duplicate Detection for Job Ads in the Online Recruitment Domain	177
<i>Hunter Burk, Faizan Javed, and Janani Balaji</i>	

Employer Industry Classification Using Job Postings	183
<i>Mahak Goindani, Qiaoling Liu, Josh Chao, and Valentin Jijkoun</i>	

Finding the Best Job Applicants for a Job Posting: A Comparison of Human Resources Search Strategies	189
<i>Christopher G. Harris</i>	

Survival Random Forest to Predict Time to Fill	195
<i>Summer M. Husband and Jason Roberts</i>	

Data-Driven Job Search Engine Using Skills and Company Attribute Filters	199
<i>Rohit Muthyala, Sam Wood, Yi Jin, Yixing Qin, Hua Gao, and Amit Rai</i>	

On Analyzing Job Hop Behavior and Talent Flow Networks	207
<i>Richard J. Oentaryo, Xavier Jayaraj Siddarth Ashok, Ee-Peng Lim, and Philips Kokoh Prasetyo</i>	
Long Tail Query Enrichment for Semantic Job Search	215
<i>Layla Pournajaf, Khalifeh Aljadda, and Mohammed Korayem</i>	
What's Next? A Recommendation System for Industrial Training	221
<i>Rajiv Srivastava, Girish Keshav Palshikar, and Saheb Chourasia</i>	
Estimating Fungibility Between Skills by Combining Skill-Similarities Obtained from Multiple Data Sources	229
<i>Shrihari Vasudevan, Moninder Singh, Joydeep Mondal, Michael Peran, Ben Zweig, Brian Johnston, and Rachel Rosenfeld</i>	
On Utility of Temporal Embeddings in Skill Matching	237
<i>Manisha Verma and Nathan Francis</i>	

SSTDM - Spatial and Spatiotemporal Data Mining

Spectral Keyboard Streams: Towards Effective and Continuous Authentication	242
<i>Abdullah Alshehri, Frans Coenen, and Danushka Bollegala</i>	
Top-(R%, K) Spatiotemporal Event Sequence Mining	250
<i>Berkay Aydin, Ahmet Kucuk, Soukaina Filali Boubrahimi, and Rafal A. Angryk</i>	
Identifying Irregular Power Usage by Turning Predictions into Holographic Spatial Visualizations	258
<i>Patrick Glauner, Niklas Dahringer, Oleksandr Puhachov, Jorge Augusto Meira, Petko Valtchev, Radu State, and Diogo Duarte</i>	
Discovering Gatherings Based on Individual Mobility Patterns: Challenges and Direction	266
<i>Fatima Hachem, Maria Luisa Damiani, and Hamza Issa</i>	
Cultivating Evolving Region Trajectory Datasets	274
<i>Sajitha Naduvil-Vadukootu, Berkay Aydin, Michael A. Schuh, and Rafal A. Angryk</i>	
Stop Purpose Classification from GPS Data of Commercial Vehicle Fleets	280
<i>Leonardo Sarti, Luca Bravi, Francesco Sambo, Leonardo Taccari, Matteo Simoncini, Samuele Salti, and Alessandro Lori</i>	
Pruning and Nonparametric Multiple Change Point Detection	288
<i>Wenyu Zhang, Nicholas A. James, and David S. Matteson</i>	

DMESS - Data Mining in Earth System Science

Quantifying Seasonal Patterns in Disparate Environmental Variables Using the PolarMetrics R Package	296
<i>Bjorn-Gustaf J. Brooks, Danny C. Lee, Lars Y. Pomara, William W. Hargrove, and Ankur R. Desai</i>	
Deriving Data-Driven Insights from Climate Extreme Indices for the Continental US	303
<i>Xinbo Huang, David Sathiaraj, Lei Wang, and Barry Keim</i>	

Resolution Reconstruction of Climate Data with Pixel Recursive Model	313
<i>Sookyung Kim, Sasha Ames, Jiwoo Lee, Chengzhu Zhang, Aaron C. Wilson, and Dean Williams</i>	
Convolutional Neural Network Approach for Mapping Arctic Vegetation Using Multi-Sensor Remote Sensing Fusion	322
<i>Zachary L. Langford, Jitendra Kumar, and Forrest M. Hoffman</i>	
A Machine Learning Approach to Non-uniform Spatial Downscaling of Climate Variables	332
<i>Soukayna Mouatadid, Steve Easterbrook, and Andre R. Erler</i>	
SENTIRE - Sentiment Elicitation from Natural Text for Information Retrieval and Extraction	
Analyzing Informal Caregiving Expression in Social Media	342
<i>Reda Al-Bahrani, Margaret K. Danilovich, Wei-Keng Liao, Alok Choudhary, and Ankit Agrawal</i>	
Estimating Personality from Social Media Posts	350
<i>Nasser Alsadhan and David Skillicorn</i>	
Learning-Based Method with Valence Shifters for Sentiment Analysis	357
<i>Ruihua Cheng and Ji Meng Loh</i>	
Twitter Stance Detection — A Subjectivity and Sentiment Polarity Inspired Two-Phase Approach	365
<i>Kuntal Dey, Ritvik Shrivastava, and Saroj Kaushik</i>	
Dataset Construction via Attention for Aspect Term Extraction with Distant Supervision	373
<i>Athanasios Giannakopoulos, Diego Antognini, Claudiu Musat, Andreea Hossmann, and Michael Baeriswyl</i>	
Analyzing Users' Sentiment Towards Popular Consumer Industries and Brands on Twitter	381
<i>Guoning Hu, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger, and Nemanja Spasojevic</i>	
An Experimental Evaluation of Prior Polarities in Sentiment Lexicons	389
<i>Ali Bugra Kanburoglu and Ercan Solak</i>	
Let's Chat about Brexit! A Politically-Sensitive Dialog System Based on Twitter Data	393
<i>Aparup Khatua, Erik Cambria, Apalak Khatua, and Iti Chaturvedi</i>	
Sentiment Extraction from Consumer-Generated Noisy Short Texts	399
<i>Hardik Meisheri, Kunal Ranjan, and Lipika Dey</i>	
Phonetic-Based Microtext Normalization for Twitter Sentiment Analysis	407
<i>Ranjan Satapathy, Claudia Guerreiro, Iti Chaturvedi, and Erik Cambria</i>	
A Bootstrap Method for Automatic Rule Acquisition on Emotion Cause Extraction	414
<i>Shuntaro Yada, Kazushi Ikeda, Keiichiro Hoashi, and Kyo Kageura</i>	
Extracting User-Reported Mobile Application Defects from Online Reviews	422
<i>Yue Wang, Hongning Wang, and Hui Fang</i>	

DMBIH - Data Mining in Biomedical Informatics and Healthcare

An CNN-LSTM Attention Approach to Understanding User Query Intent from Online Health Communities	430
<i>Ruichu Cai, Binjun Zhu, Lei Ji, Tianyong Hao, Jun Yan, and Wenyin Liu</i>	
Process-Oriented Iterative Multiple Alignment for Medical Process Mining	438
<i>Shuhong Chen, Sen Yang, Moliang Zhou, Randall Burd, and Ivan Marsic</i>	
Exploiting PubMed for Protein Molecular Function Prediction via NMF Based Multi-label Classification	446
<i>Samah Fodeh, Aditya Tiwari, and Hong Yu</i>	
Discovery of Informal Topics from Post Traumatic Stress Disorder Forums	452
<i>Reilly Grant, David Kucher, Ana M. León, Jonathan Gemmell, and Daniela Raicu</i>	
RESTRAC: REference Sequence Based Space TRAnsformation for Clustering	462
<i>AKM Tauhidul Islam, Sakti Pramanik, Vahid Mirjalili, and Shamik Sural</i>	
Probable Biomarker Identification Using Recursive Feature Extraction and Network Analysis	470
<i>Arpit Mishra, Abhishek Gupta, Umesh Maheswari, and Laeeq Siddique</i>	
Semi-Supervised Prediction of Comorbid Rare Conditions Using Medical Claims Data	478
<i>Chirag Nagpal, Kyle Miller, Tiffany Pellathy, Marilyn Hravnak, Gilles Clermont, Michael Pinsky, and Artur Dubrawski</i>	
Deep Physiological Arousal Detection in a Driving Simulator Using Wearable Sensors	486
<i>Aaqib Saeed, Stojan Trajanovski, Maurice van Keulen, and Jan van Erp</i>	
GB-R: A Fast and Effective Gray-Box Reconstruction of Cascade Time-Series	494
<i>Hyun Ah Song, Fan Yang, Zongge Liu, Wilbert van Panhuis, Nicholas Sidiropoulos, Christos Faloutsos, and Vladimir Zadorozhny</i>	
Detecting Opioid Users from Twitter and Understanding Their Perceptions Toward MAT	502
<i>Yiming Zhang, Yujie Fan, Yanfang Ye, Xin Li, and Wanhong Zheng</i>	

HDM - High Dimensional Data Mining

Robust Projective Dictionary Learning by Joint Label Embedding and Classification	510
<i>Weiming Jiang, Zhao Zhang, Jie Qin, Mingbo Zhao, Fanzhang Li, and Shuicheng Yan</i>	
Taming Wild High Dimensional Text Data with a Fuzzy Lash	518
<i>Amir Karami</i>	
High-Dimensional Density Estimation for Data Mining Tasks	523
<i>Alexander Kuleshov, Alexander Bernstein, and Yury Yanovich</i>	
Multiple Queries of Information Retrieval Using Krylov Subspace Method	531
<i>Youzuo Lin</i>	
Differential Geometric Retrieval of Deep Features	539
<i>Y. Qian, E. Vazquez, and B. Sengupta</i>	
Evaluation of Non-linearity in MIR Spectroscopic Data for Compressed Learning	545
<i>Dixon Vimalajeewa, Donagh Berry, Eric Robson, and Chamil Kulatunga</i>	

Near-Optimal Noisy Low-Tubal-Rank Tensor Completion via Singular Tube Thresholding	553
<i>Andong Wang and Zhong Jin</i>	
Lazy Stochastic Principal Component Analysis	561
<i>Michael Wojnowicz, Dinh Nguyen, Li Li, and Xuan Zhao</i>	
A Novel Method for Fast and Accurate Similarity Measure in Time Series Field	569
<i>Miaomiao Zhang and Dechang Pi</i>	

MoDat - Market of Data: Creating tools, data, and sensors from the Social Intelligence

VARIABLE QUEST: Network Visualization of Variable Labels Unifying Co-occurrence Graphs	577
<i>Teruaki Hayashi and Yukio Ohsawa</i>	
Development and Use of an Activity Model Based on Structured Knowledge: A Music Teaching Support System	584
<i>Nami Iino, Satoshi Nishimura, Ken Fukuda, Kentaro Watanabe, Jokinen Kristiina, and Takuichi Nishimura</i>	
Extracting Field Overseers' Features in Risk Recognition from Data of Eyes and Utterances	590
<i>Noriyuki Kushiro, Yuji Fujita, and Yusuke Aoyama</i>	
Familiarity and Strangeness of Objects: A MoDAT Requirement for Shikake Design	597
<i>Naohiro Matsumura</i>	
Semantic Visualization Support for Innovators Marketplace on Data Jackets	599
<i>Qi Wang</i>	

DMS - Data Mining for Service

Global Distribution of Watches: A Network Analysis of Trade Relations	605
<i>Pierre-Yves Donzé, Ken Ishibashi, Bo Wu, Yuta Kaneko, Kei Miyazaki, and Keiji Takai</i>	
Frequent Temporal Pattern Mining for Medical Data Based on Ranged Relations	612
<i>Shoji Hirano and Shusaku Tsumoto</i>	
Personal Identification by Pedestrians Behavior	617
<i>Eisuke Kita, Xuanang Feng, and Hiroki Shimokubo</i>	
Forecasting of Commercial Sales with Large Scale Gaussian Processes	625
<i>Rodrigo Rivera and Evgeny Burnaev</i>	
Akori: A Tool Based in Eye-Tracking Techniques for Analyzing Web User Behaviour on a Web Site	635
<i>Felipe Vera, Víctor D. Cortés, Gabriel Iturra, Juan D. Velásquez, Pedro Maldonado, and Andrés Couve</i>	
Social Intimacy Based IoT Services Mining of Massive Data	641
<i>Anni Zhou, Yinan Feng, Pan Zhou, and Jie Xu</i>	

DMCS - Data Mining for Cyber Security

Data Mining in Long-Term Honeypot Data	649
<i>Daniel Fraunholz, Marc Zimmermann, Alexander Hafner, and Hans D. Schotten</i>	
IP2Vec: Learning Similarities Between IP Addresses	657
<i>Markus Ring, Alexander Dallmann, Dieter Landes, and Andreas Hotho</i>	
Early Warnings of Cyber Threats in Online Discussions	667
<i>Anna Sapienza, Alessandro Bessi, Saranya Damodaran, Paulo Shakarian, Kristina Lerman, and Emilio Ferrara</i>	
Meta-Morisita Index: Anomaly Behaviour Detection for Large Scale Tracking Data with Spatio-Temporal Marks	675
<i>Zhao Yang and Nathalie Japkowicz</i>	
Inline DGA Detection with Deep Networks	683
<i>Bin Yu, Daniel L. Gray, Jie Pan, Martine De Cock, and Anderson C. A. Nascimento</i>	

ARIAL - Data mining for Aging, Rehabilitation and Assisted Living

Towards Automatic Feature Extraction for Activity Recognition from Wearable Sensors: A Deep Learning Approach	693
<i>Belkacem Chikhaoui and Frank Gouineau</i>	
DAAD: A Framework for Detecting Agitation and Aggression in People Living with Dementia Using a Novel Multi-modal Sensor Network	703
<i>Shehroz S. Khan, Tong Zhu, Bing Ye, Alex Mihailidis, Andrea Iaboni, Kristine Newman, Angel He Wang, and Lori Schindel Martin</i>	
A Virtual Research Environment to Support Remote Behaviour Monitoring from Data Collection to Analysis	711
<i>Frank Krüger, Sebastian Bader, Albert Hein, and Thomas Kirste</i>	
Predicting Dementia Screening and Staging Scores from Semantic Verbal Fluency Performance	719
<i>Nicklas Linz, Johannes Tröger, Jan Alexandersson, Maria Wolters, Alexandra König, and Philippe Robert</i>	
Detection of Residents' Abnormal Behaviour by Analysing Energy Consumption of Individual Households	729
<i>Christian Nordahl, Marie Persson, and Håkan Grahm</i>	
Feature Selection for the Classification of Longitudinal Human Ageing Data	739
<i>Tossapol Pomsuwan and Alex A. Freitas</i>	

SERecSys - Semantics-Enabled Recommender Systems

Semantic Search-by-Examples for Scientific Topic Corpus Expansion in Digital Libraries	747
<i>Hussein T. Al-Natsheh, Lucie Martinet, Fabrice Muhlenbach, Fabien Rico, and Djamel A. Zighed</i>	
Leveraging Moderate User Data for News Recommendation	757
<i>Dhruv Khattar, Vaibhav Kumar, and Vasudeva Varma</i>	
Word Semantics Based 3-D Convolutional Neural Networks for News Recommendation	761
<i>Vaibhav Kumar, Dhruv Khattar, Shashank Gupta, and Vasudeva Varma</i>	
User Profiling Based Deep Neural Network for Temporal News Recommendation	765
<i>Vaibhav Kumar, Dhruv Khattar, Shashank Gupta, Manish Gupta, and Vasudeva Varma</i>	
Sequential Heterogeneous Attribute Embedding for Item Recommendation	773
<i>Kuan Liu, Xing Shi, and Prem Natarajan</i>	
Entity Recommendation Via Integrating Multiple Types of Implicit Feedback in Heterogeneous Information Network	781
<i>Xiaotong Suo, Fang Wei, and Ke Yu</i>	

ISI - Intelligence and Security Informatics

Finding Suspicious Activities in Financial Transactions and Distributed Ledgers	787
<i>Ramiro Daniel Camino, Radu State, Leandro Montero, and Petko Valtchev</i>	
Anomalous User Activity Detection in Enterprise Multi-source Logs	797
<i>Qiaona (Joanna) Hu, Baoming Tang, and Derek Lin</i>	
Reducing False Positives of User-to-Entity First-Access Alerts for User Behavior Analytics	804
<i>Baoming Tang, Qiaona (Joanna) Hu, and Derek Lin</i>	

OEDM - Optimization Based Techniques for Emerging Data Mining

A Hybrid Training Algorithm for Recurrent Neural Network Using Particle Swarm Optimization-Based Preprocessing and Temporal Error Aggregation	812
<i>Qiao Kang, Wei-Keng Liao, Ankit Agrawal, and Alok Choudhary</i>	
A New Method for Stock Price Prediction Based on MRFs and SSVM	818
<i>Lin Lai, Chang Li, and Wen Long</i>	
A Feasible Direction Method for Optimization Problem with Orthogonal Constraint in Feature Selection	824
<i>Jianyu Miao, Yong Shi, and Lingfeng Niu</i>	
A Novel l_0 -Constrained Gaussian Graphical Model for Anomaly Localization	830
<i>Dzung T. Phan, Tsuyoshi Idé, Jayant Kalagnanam, Matt Menickelly, and Katya Scheinberg</i>	

DMIP - Data Mining in Practice: Automation and Cost

An Empirical Evaluation of Techniques for Feature Selection with Cost	834
<i>Stephen Adams, Ryan Meekins, and Peter A. Beling</i>	
The Mean and Median Criteria for Kernel Bandwidth Selection for Support Vector Data Description	842
<i>Arin Chaudhuri, Deovrat Kakde, Carol Sadek, Laura Gonzalez, and Seunghyun Kong</i>	
Combining Active Learning and Semi-Supervised Learning by Using Selective Label Spreading	850
<i>Xu Chen and Tao Wang</i>	
Robust Self-Tuning Sparse Subspace Clustering	858
<i>Guangtao Wang, Jiayu Zhou, Jingjie Ni, Tingjin Luo, Wei Long, Hai Zhen, Gao Cong, and Jieping Ye</i>	

DMiP - Data Mining in Politics

National Leaders' Twitter Speech to Infer Political Leaning and Election Results in 2015 Venezuelan Parliamentary Elections	866
<i>Rodrigo Castro and Carmen Vaca</i>	
Us and Them: Adversarial Politics on Twitter	872
<i>Anna Guimarães, Liqiang Wang, and Gerhard Weikum</i>	
Measuring Network Structure Metrics as a Proxy for Socio-Political Activity in Social Media	878
<i>Selvas Mwanza and Hussein Suleman</i>	
Controversy Detection Using Reactions on Social Media	884
<i>Allaparthi Sriteja, Prakhar Pandey, and Vikram Pudi</i>	

DAPS - Data Mining for the Analysis of Performance and Success

GPS Data Reflect Players' Internal Load in Soccer	890
<i>Alessio Rossi, Enrico Perri, Athos Trecroci, Marco Savino, Giampietro Alberti, and F Marcello Iaia</i>	
Identifying Basketball Plays from Sensor Data; Towards a Low-Cost Automatic Extraction of Advanced Statistics	894
<i>Adrià Arbués Sangüesa, Thomas B. Moeslund, Chris H. Bahnsen, and Raul Benítez Iglesias</i>	
Performance Dynamics and Success in Online Games	902
<i>Anna Sapienza, Hao Peng, and Emilio Ferrara</i>	

BDDSR - Workshop on Big Data & Data Science in Retail

Mobile E-Commerce Data Processing Using Relational Memory	910
<i>Parham Aarabi</i>	
A Pattern Tree Based Method for Mining Conditional Contrast Patterns of Multi-source Data	916
<i>Li Li, Sarah Erfani, and Christopher Leckie</i>	

Improving Multivariate Time Series Forecasting with Random Walks with Restarts on Causality Graphs	924
<i>Piotr Przymus, Youssef Hmamouche, Alain Casali, and Lotfi Lakhal</i>	

D3M - Data-driven Discovery of Models

Dataset Selection for Controlling Swarms by Visual Demonstration	932
<i>Karan Kumar Budhraj and Tim Oates</i>	
Discrimination at the Edge of Noise: A Hilbert Space of Stationary Ergodic Processes	942
<i>Ishanu Chattopadhyay</i>	
Ranking from Crowdsourced Pairwise Comparisons via Smoothed Matrix Manifold Optimization	949
<i>Jialin Dong, Kai Yang, and Yuanming Shi</i>	
Feature Selection in Learning Using Privileged Information	957
<i>Rauf Izmailov, Blerta Lindqvist, and Peter Lin</i>	
A Central Limit Theorem for an Omnibus Embedding of Multiple Random Dot Product Graphs	964
<i>Keith Levin, Avanti Athreya, Minh Tang, Vince Lyzinski, and Carey E. Priebe</i>	

HPGDML - High Performance Graph Data Mining and Machine Learning Workshop

Online Detection of Anomalous Heterogeneous Graphs with Streaming Edges	968
<i>Leman Akoglu</i>	
Review of Graph Processing Frameworks	N/A
<i>Siddharth Bhatia and Rajiv Kumar</i>	
High Performance Graph Data Management and Mining with X10	977
<i>Miyuru Dayarathna</i>	
Inferring, Summarizing and Mining Multi-source Graph Data	978
<i>Danai Koutra</i>	
Image Analysis Using Convolutional Neural Networks for Modeling 2D Fracture Propagation	979
<i>Robyn L. Miller, Bryan Moore, Hari Viswanathan, and Gowri Srinivasan</i>	
Inside the Atoms: Mining a Network of Networks and Beyond	983
<i>Hanghang Tong</i>	

PAIS - Privacy and Anonymity in the Information Society

Factor Analysis for Anonymization	984
<i>Aida Calviño, Palmira Aldeguer, and Josep Domingo-Ferrer</i>	
Privacy-Preserving Big Data Stream Mining: Opportunities, Challenges, Directions	992
<i>Alfredo Cuzzocrea</i>	
Steered Microaggregation: A Unified Primitive for Anonymization of Data Sets and Data Streams	995
<i>Josep Domingo-Ferrer and Jordi Soria-Comas</i>	

Personalized Anonymization for Set-Valued Data by Partial Suppression	1003
<i>Takuma Nakagawa, Hiromi Arai, and Hiroshi Nakagawa</i>	
Revealing the Unseen: How to Expose Cloud Usage While Protecting User Privacy	1011
<i>Ata Turk, Mayank Varia, and Georgios Kellaris</i>	

ACUMEN - Data Science for Human Performance in Social Networks

Co-Training for Demographic Classification Using Deep Learning from Label Proportions	1017
<i>Ehsan Mohammady Ardehaly and Aron Culotta</i>	
EmTaggeR: A Word Embedding Based Novel Method for Hashtag Recommendation on Twitter	1025
<i>Kuntal Dey, Ritvik Shrivastava, Saroj Kaushik, and L. Venkata Subramaniam</i>	
Discovering Cooperative Structure Among Online Items for Attention Dynamics	1033
<i>Kanji Matsutani, Masahito Kumano, Masahiro Kimura, Kazumi Saito, Kouzou Ohara, and Hiroshi Motoda</i>	
Live on TV, Alive on Twitter: Quantifying Continuous Partial Attention of Viewers During Live Television Telecasts	1042
<i>Rohit Saxena, Savita Bhat, and Niranjan Pedanekar</i>	
Are Words Commensurate with Actions? Quantifying Commitment to a Cause from Online Public Messaging	1050
<i>Zhao Wang, Jennifer Cutler, and Aron Culotta</i>	

DMCIS - Data Mining for Cyberphysical and Industrial Systems

Anomaly Detection for a Water Treatment System Using Unsupervised Machine Learning	1058
<i>Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M. Poskitt, and Jun Sun</i>	
Pattern-Based Contextual Anomaly Detection in HVAC Systems	1066
<i>Mohsin Munir, Steffen Erkel, Andreas Dengel, and Sheraz Ahmed</i>	
An Adaptive Modeling Framework for Bivariate Data Streams with Applications to Change Detection in Cyber-Physical Systems	1074
<i>Joshua Plasse, Jordan Noble, and Kary Myers</i>	
Data-Driven Anomaly Detection for Power System Generation Control	1082
<i>Pengyuan Wang, Manimaran Govindarasu, Aditya Ashok, Siddharth Sridhar, and David McKinnon</i>	

PhD Forum

Dependency Anomaly Detection for Heterogeneous Time Series: A Granger-Lasso Approach	1090
<i>Sahar Behzadi, Katerina Hlaváčková-Schindler, and Claudia Plant</i>	
Deep Learning Solutions to Computational Phenotyping in Health Care	1100
<i>Zhengping Che and Yan Liu</i>	

Aggregation and Disaggregation of Information: A Holistic View	1110
<i>Yuyue Chen and Chuanren Liu</i>	
Deep Learning for Pulmonary Nodule CT Image Retrieval — An Online Assistance System for Novice Radiologists	1112
<i>Daniel Perez Ibanez, Jiang Li, Yuzhong Shen, Joan Dayanghirang, Shengli Wang, and Zezhong Zheng</i>	
Development of an Interpretable Neural Network Model for Creation of Polarity Concept Dictionaries	1122
<i>Tomoki Ito, Hiroki Sakaji, Kiyoshi Izumi, Kota Tsubouchi, and Tatsuo Yamashita</i>	
Uncovering Teamwork in Networks — Prediction, Optimization and Explanation	1132
<i>Liangyue Li and Hanghang Tong</i>	
A Multilevel NER Framework for Automatic Clinical Name Entity Recognition	1134
<i>Thoai Man Luu, Robert Phan, Rachel Davey, and Girija Chetty</i>	
Network Embedding with Centrality Information	1144
<i>Yao Ma, Suhang Wang, and Jiliang Tang</i>	
Predicting Weight Loss with Ensemble Methods	1146
<i>Zhiwei Wang and Jiliang Tang</i>	
Crowdsourcing Data Science for Innovation	1148
<i>Wangcheng Yan, Wenjun Zhou, Paolo Letizia, and Bogdan Bichescu</i>	
Exploring Transfer Learning for Crime Prediction	1158
<i>Xiangyu Zhao and Jiliang Tang</i>	

Demo Session

FRS: Fast Range Search by Pruning Unnecessary Distance Computations Based on K-D Tree	1160
<i>Yewang Chen, Jai Puneet Singh, Lida Zhou, and Nizar Bouguila</i>	
Pi-CEP: Predictive Complex Event Processing Using Range Queries over Historical Pattern Space	1166
<i>Syed Gillani, Abderrahmen Kammoun, Kamal Singh, Julien Subercaze, Christophe Gravier, Jacques Fayolle, and Frédérique Laforest</i>	
Mining Human Mobility to Quantify Performance Status	1172
<i>Minh N.B. Nguyen, Zaki Hasnain, Ming Li, Tanya Dorff, David Quinn, Sanjay Purushotham, Luciano Nocera, Paul K. Newton, Peter Kuhn, Jorge Nieva, and Cyrus Shahabi</i>	
Tweeloc: A System for Geolocating Tweets at Fine-Grain	1178
<i>Pavlos Paraskevopoulos, Giovanni Pellegrini, and Themis Palpanas</i>	

Additional Paper

A Budget-Constrained Inverse Classification Framework for Smooth Classifiers.....	1184
<i>Michael T. Lash, Qihang Lin, W. Nick Street and Jennifer G. Robinson</i>	

Author Index