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
  Biswarup Bhattacharya, Iftikhar Burhanuddin, Abhilasha Sancheti, and Kushal Satya
A Big Data Analysis Framework Using Apache Spark and Deep Learning .................................................................9
  Anand Gupta, Hardeo Kumar Thakur, Ritvik Shrivastava, Pulkit Kumar, and Sreyashi Nag
Analyzing Dynamical Activities of Co-occurrence Patterns for Cooking Ingredients ..............................................17
  Yuuki Kikuchi, Masahito Kumano, and Masahiro Kimura
Distance and Density Clustering for Time Series Data .................................................................................................25
  Ruizhe Ma and Rafal Angryk
Accelerated Hierarchical Density Based Clustering .....................................................................................................33
  Leland McInnes and John Healy
Hybrid.poly: An Interactive Large-Scale In-memory Analytical Polystore ....................................................................43
  Maksim Podkorytov, Dylan Soderman, and Michael Gubanov
Action Rules for Sentiment Analysis on Twitter Data Using Spark ........................................................................51
  Jaishree Ranganathan, Allen S. Irudayaraj, and Angelina A. Tzacheva
Automated Storytelling Evaluation and Story Chain Generation ..................................................................................61
  J.T. Rigsby and Daniel Barbarà
Dealing with Class Imbalance the Scalable Way: Evaluation of Various Techniques Based on Classification Grade and Computational Complexity .................................................................69
  Bernhard Schlegel and Bernhard Sick
Failure Prediction with Adaptive Multi-scale Sampling and Activation Pattern Regularization ..............................................79
  Yujin Tang, Shinya Wada, and Kiyohito Yoshihara
Discovery of Action Rules at Lowest Cost in Spark ................................................................................................87
  Angelina A. Tzacheva, Arunkumar Bagavathi, and Lavanya Ayila
Estimating Treatment Effects of a Residential Demand Response Program Using Non-experimental Data ..................................................................................................................95
Datong Paul Zhou, Maximilian Balandat, and Claire Jennifer Tomlin

Deep and Confident Prediction for Time Series at Uber ..........................................................103
Lingxue Zhu and Nikolay Laptev

DaMNet - Data Mining in Networks
Distributed Representations of Subgraphs ..................................................................................111
Bijaya Adhikari, Yao Zhang, Naren Ramakrishnan, and B. Aditya Prakash

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

Mining Patterns of Cyberbullying on Twitter ...........................................................................126
Charalampos Chelmis, Daphney-Stavroula Zois, and Mengfan Yao

Principled Multilayer Network Embedding .................................................................................134
Weiyi Liu, Pin-yu Chen, Sailung Yeung, Toyotaro Suzumura, and Lingli Chen

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

Predicting Users’ Search Behavior Using Stochastic Multi-mode Network Models .................150
Shohei Umehara and Koji Eguchi

Constructing a Hierarchical User Interest Structure Based on User Profiles .............................156
Chao Zhao, Min Zhao, and Yi Guan

Exploring Uncertainty Methods for Centrality Analysis in Social Networks ............................163
Xianglin Zuo, Bo Yang, and Wanli Zuo

DSHCM - Data Science for Human Capital Management
Analysis of Organizational Structure Through Cluster Validation Techniques: Evaluation of Email Communications at an Organizational Level .................................................................170
Veselka Boeva, Lars Lundberg, Sai M. Harsha Kota, and Lars Sköld

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

Employer Industry Classification Using Job Postings .................................................................183
Mahak Goindani, Qiaoling Liu, Josh Chao, and Valentin Jijkoun

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

Survival Random Forest to Predict Time to Fill .........................................................................195
Summer M. Husband and Jason Roberts

Data-Driven Job Search Engine Using Skills and Company Attribute Filters ............................199
Rohit Muthyala, Sam Wood, Yi Jin, Yixing Qin, Hua Gao, and Amit Rai
On Analyzing Job Hop Behavior and Talent Flow Networks ................................................................. 207
   Richard J. Oentaryo, Xavier Jayaraj Siddarth Ashok, Ee-Peng Lim, and Philips Kokoh Prasetyo

Long Tail Query Enrichment for Semantic Job Search ........................................................................ 215
   Layla Pournajaf, Khalifeh Aljadda, and Mohammed Korayem

What's Next? A Recommendation System for Industrial Training ......................................................... 221
   Rajiv Srivastava, Girish Keshav Palshikar, and Saheb Chourasia

Estimating Fungibility Between Skills by Combining Skill-Similarities Obtained from Multiple Data Sources ........................................................................................................... 229
   Shrihari Vasudevan, Moninder Singh, Joydeep Mondal, Michael Peran, Ben Zweig, Brian Johnston, and Rachel Rosenfeld

On Utility of Temporal Embeddings in Skill Matching ......................................................................... 237
   Manisha Verma and Nathan Francis

SSTDM - Spatial and Spatiotemporal Data Mining

Spectral Keyboard Streams: Towards Effective and Continuous Authentication .................................... 242
   Abdullah Alshehri, Frans Coenen, and Danushka Bollegala

Top-(R%, K) Spatiotemporal Event Sequence Mining ........................................................................... 250
   Berkay Aydin, Ahmet Kucuk, Soukaina Filali Boubrahimi, and Rafal A. Angryk

Identifying Irregular Power Usage by Turning Predictions into Holographic Spatial Visualizations ........ 258
   Patrick Glauner, Niklas Dahringer, Oleksandr Puhachov, Jorge Augusto Meira, Petko Valtchev, Radu State, and Diogo Duarte

Discovering Gatherings Based on Individual Mobility Patterns: Challenges and Direction ................... 266
   Fatima Hachem, Maria Luisa Damiani, and Hamza Issa

Cultivating Evolving Region Trajectory Datasets .................................................................................... 274
   Sajitha Naduvil-Vadukootu, Berkay Aydin, Michael A. Schuh, and Rafal A. Angryk

Stop Purpose Classification from GPS Data of Commercial Vehicle Fleets ........................................... 280
   Leonardo Sarti, Luca Bravi, Francesco Sambo, Leonardo Taccari, Matteo Simoncini, Samuele Salti, and Alessandro Lori

Pruning and Nonparametric Multiple Change Point Detection ............................................................... 288
   Wenyu Zhang, Nicholas A. James, and David S. Matteson

DMESS - Data Mining in Earth System Science

Quantifying Seasonal Patterns in Disparate Environmental Variables Using the PolarMetrics R Package ............................................................................................................................ 296

Deriving Data-Driven Insights from Climate Extreme Indices for the Continental US .......................... 303
   Xinbo Huang, David Sathiaraj, Lei Wang, and Barry Keim
Resolution Reconstruction of Climate Data with Pixel Recursive Model .................................................................313
    Sookyung Kim, Sasha Ames, Jiwoo Lee, Chengzhu Zhang, Aaron C. Wilson, and Dean Williams

Convolutional Neural Network Approach for Mapping Arctic Vegetation Using Multi-Sensor Remote Sensing Fusion .................................................................322
    Zachary L. Langford, Jitendra Kumar, and Forrest M. Hoffman

A Machine Learning Approach to Non-uniform Spatial Downscaling of Climate Variables .................................................................332
    Soukayna Mouatadid, Steve Easterbrook, and Andre R. Erler

SENTIRE - Sentiment Elicitation from Natural Text for Information Retrieval and Extraction

Analyzing Informal Caregiving Expression in Social Media .........................................................................................342
    Reda Al-Bahrani, Margaret K. Danilovich, Wei-Keng Liao, Alok Choudhary, and Ankit Agrawal

Estimating Personality from Social Media Posts .........................................................................................................350
    Nasser Alsadhan and David Skillicorn

Learning-Based Method with Valence Shifters for Sentiment Analysis ..........................................................................357
    Ruihua Cheng and Ji Meng Loh

Twitter Stance Detection — A Subjectivity and Sentiment Polarity Inspired Two-Phase Approach .................................................................365
    Kuntal Dey, Ritvik Shrivastava, and Saroj Kaushik

Dataset Construction via Attention for Aspect Term Extraction with Distant Supervision .............................................373
    Athanasios Giannakopoulos, Diego Antognini, Claudiu Musat, Andreea Hossmann, and Michael Baeriswyl

Analyzing Users’ Sentiment Towards Popular Consumer Industries and Brands on Twitter .........................................................................................381
    Guoning Hu, Preeti Bhargava, Saul Fuhrmann, Sarah Ellinger, and Nemanja Spasojevic

An Experimental Evaluation of Prior Polarities in Sentiment Lexicons ........................................................................389
    Ali Bugra Kanburoglu and Ercan Solak

Let’s Chat about Brexit! A Politically-Sensitive Dialog System Based on Twitter Data .................................................393
    Aparup Khatua, Erik Cambria, Apalak Khatua, and Iti Chaturvedi

Sentiment Extraction from Consumer-Generated Noisy Short Texts ........................................................................399
    Hardik Meisher, Kunal Ranjan, and Lipika Dey

Phonetic-Based Microtext Normalization for Twitter Sentiment Analysis ........................................................................407
    Ranjan Satapathy, Claudia Guerreiro, Iti Chaturvedi, and Erik Cambria

A Bootstrap Method for Automatic Rule Acquisition on Emotion Cause Extraction .................................................414
    Shuntaro Yada, Kazushi Ikeda, Keichiro Hoashi, and Kyo Kageura

Extracting User-Reported Mobile Application Defects from Online Reviews .................................................................422
    Yue Wang, Hongning Wang, and Hui Fang
DMBIH - Data Mining in Biomedical Informatics and Healthcare

An CNN-LSTM Attention Approach to Understanding User Query Intent from Online Health Communities .................................................................430
Ruichu Cai, Binjun Zhu, Lei Ji, Tianyong Hao, Jun Yan, and Wenjin Liu

Process-Oriented Iterative Multiple Alignment for Medical Process Mining .................................................................438
Shuhong Chen, Sen Yang, Moliang Zhou, Randall Burd, and Ivan Marsic

Exploiting PubMed for Protein Molecular Function Prediction via NMF Based Multi-label Classification .................................................................446
Samah Fodeh, Aditya Tiwari, and Hong Yu

Discovery of Informal Topics from Post Traumatic Stress Disorder Forums .................................................................452
Reilly Grant, David Kucher, Ana M. León, Jonathan Gemmell, and Daniela Raicu

RESTRAC: REference Sequence Based Space TRAnsformation for Clustering .................................................................462
AKM Tauhidul Islam, Sakti Pramanik, Vahid Mirjalili, and Shamik Sural

Probable Biomarker Identification Using Recursive Feature Extraction and Network Analysis .................................................................470
Arpit Mishra, Abhishek Gupta, Umesh Maheswari, and Laeeq Siddique

Semi-Supervised Prediction of Comorbid Rare Conditions Using Medical Claims Data .................................................................478
Chirag Nagpal, Kyle Miller, Tiffany Pellathy, Marilyn Hravnak, Gilles Clermont, Michael Pinsky, and Artur Dubrawski

Deep Physiological Arousal Detection in a Driving Simulator Using Wearable Sensors .................................................................486
Aaqib Saeed, Stojan Trajanovski, Maurice van Keulen, and Jan van Erp

GB-R: A Fast and Effective Gray-Box Reconstruction of Cascade Time-Series .................................................................494
Hyun Ah Song, Fan Yang, Zongge Liu, Wilbert van Panhuis, Nicholas Sidiropoulos, Christos Faloutsos, and Vladimir Zadorozhny

Detecting Opioid Users from Twitter and Understanding Their Perceptions Toward MAT .................................................................502
Yiming Zhang, Yujie Fan, Yanfang Ye, Xin Li, and Wanhong Zheng

HDM - High Dimensional Data Mining

Robust Projective Dictionary Learning by Joint Label Embedding and Classification .................................................................510
Weiming Jiang, Zhao Zhang, Jie Qin, Mingbo Zhao, Fanzhang Li, and Shuicheng Yan

Taming Wild High Dimensional Text Data with a Fuzzy Lash .................................................................518
Amir Karami

High-Dimensional Density Estimation for Data Mining Tasks .................................................................523
Alexander Kuleshov, Alexander Bernstein, and Yury Yanovich

Multiple Queries of Information Retrieval Using Krylov Subspace Method .................................................................531
Youzuo Lin

Differential Geometric Retrieval of Deep Features .................................................................539
Y. Qian, E. Vazquez, and B. Sengupta

Evaluation of Non-linearity in MIR Spectroscopic Data for Compressed Learning .................................................................545
Dixon Vimalajeewa, Donagh Berry, Eric Robson, and Chamil Kulatunga
Near-Optimal Noisy Low-Tubal-Rank Tensor Completion via Singular Tube
Andong Wang and Zhong Jin .......................................................................................................................553

Lazy Stochastic Principal Component Analysis ...........................................................................................................561
Michael Wojnowicz, Dinh Nguyen, Li Li, and Xuan Zhao

A Novel Method for Fast and Accurate Similarity Measure in Time Series Field ........................................................569
Miaomiao Zhang and Dechang Pi

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
Teruaki Hayashi and Yukio Ohsawa

Development and Use of an Activity Model Based on Structured Knowledge: A Music Teaching Support System ...........................................................................................................................................584
Nami Iino, Satoshi Nishimura, Ken Fukuda, Kentaro Watanabe, Jokinen Kristina, and Takuichi Nishimura

Extracting Field Oversees’ Features in Risk Recognition from Data of Eyes and Utterances ............................................................................................................................................................590
Noriyuki Kushiro, Yuji Fujita, and Yusuke Aoyama

Familiarity and Strangeness of Objects: A MoDAT Requirement for Shikake Design ................................................597
Naohiro Matsumura

Semantic Visualization Support for Innovators Marketplace on Data Jackets ............................................................599
Qi Wang

DMS - Data Mining for Service

Global Distribution of Watches: A Network Analysis of Trade Relations .................................................................605
Pierre-Yves Donzé, Ken Ishibashi, Bo Wu, Yuta Kaneko, Kei Miyazaki, and Keiji Takai

Frequent Temporal Pattern Mining for Medical Data Based on Ranged Relations .....................................................612
Shoji Hirano and Shusaku Tsumoto

Personal Identification by Pedestrians Behavior .........................................................................................................617
Eisuke Kita, Xuanang Feng, and Hiroki Shimokubo

Forecasting of Commercial Sales with Large Scale Gaussian Processes ..................................................................625
Rodrigo Rivera and Evgeny Bumaev

Akori: A Tool Based in Eye-Tracking Techniques for Analyzing Web User Behaviour on a Web Site ..............................................................................................................................................................635
Felipe Vera, Víctor D. Cortés, Gabriel Iturra, Juan D. Velásquez, Pedro Maldonado, and Andrés Couve

Social Intimacy Based IoT Services Mining of Massive Data ......................................................................................641
Anni Zhou, Yinan Feng, Pan Zhou, and Jie Xu
DMCS - Data Mining for Cyber Security

Data Mining in Long-Term Honeypot Data .......................................................... 649
  Daniel Fraunholz, Marc Zimmermann, Alexander Hafner, and Hans D. Schotten

IP2Vec: Learning Similarities Between IP Addresses ........................................... 657
  Markus Ring, Alexander Dallmann, Dieter Landes, and Andreas Hotho

Early Warnings of Cyber Threats in Online Discussions .................................... 667
  Anna Sapienza, Alessandro Bessi, Saranya Damodaran, Paulo Shakarian,
  Kristina Lerman, and Emilio Ferrara

Meta-Morisita Index: Anomaly Behaviour Detection for Large Scale Tracking Data
with Spatio-Temporal Marks .................................................................................. 675
  Zhao Yang and Nathalie Japkowicz

Inline DGA Detection with Deep Networks ......................................................... 683
  Bin Yu, Daniel L. Gray, Jie Pan, Martine De Cock, and Anderson C. A. Nascimento

ARIAL - Data mining for Aging, Rehabilitation and Assisted Living

Towards Automatic Feature Extraction for Activity Recognition from Wearable Sensors:
A Deep Learning Approach ................................................................................... 693
  Belkacem Chikhaoui and Frank Gouineau

DAAD: A Framework for Detecting Agitation and Aggression in People Living
with Dementia Using a Novel Multi-modal Sensor Network .................................. 703
  Shehroz S. Khan, Tong Zhu, Bing Ye, Alex Mihailidis, Andrea Iaboni, Kristine Newman,
  Angel He Wang, and Lori Schindel Martin

A Virtual Research Environment to Support Remote Behaviour Monitoring from Data
Collection to Analysis ............................................................................................ 711
  Frank Krüger, Sebastian Bader, Albert Hein, and Thomas Kirste

Predicting Dementia Screening and Staging Scores from Semantic Verbal Fluency
Performance ........................................................................................................... 719
  Nicklas Linz, Johannes Tröger, Jan Alexandersson, Maria Wolters, Alexandra König,
  and Philippe Robert

Detection of Residents’ Abnormal Behaviour by Analysing Energy Consumption
of Individual Households ....................................................................................... 729
  Christian Nordahl, Marie Persson, and Håkan Grahn

Feature Selection for the Classification of Longitudinal Human Ageing Data .......... 739
  Tossapol Pomsuwan and Alex A. Freitas
SERecSys - Semantics-Enabled Recommender Systems
Semantic Search-by-Examples for Scientific Topic Corpus Expansion in Digital Libraries .......................................................................................................................................................................747
Hussein T. Al-Natsheh, Lucie Martinet, Fabrice Muhlenbach, Fabien Rico, and Djamel A. Zighed
Leveraging Moderate User Data for News Recommendation .................................................................757
Dhruv Khattar, Vaibhav Kumar, and Vasudeva Varma
Word Semantics Based 3-D Convolutional Neural Networks for News Recommendation .................................761
Vaibhav Kumar, Dhruv Khattar, Shashank Gupta, and Vasudeva Varma
User Profiling Based Deep Neural Network for Temporal News Recommendation ........................................765
Vaibhav Kumar, Dhruv Khattar, Shashank Gupta, Manish Gupta, and Vasudeva Varma
Sequential Heterogeneous Attribute Embedding for Item Recommendation ...............................................773
Kuan Liu, Xing Shi, and Prem Natarajan
Entity Recommendation Via Integrating Multiple Types of Implicit Feedback in Heterogeneous Information Network .......................................................................................................................781
Xiaotong Suo, Fang Wei, and Ke Yu

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

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
Qiao Kang, Wei-Keng Liao, Ankit Agrawal, and Alok Choudhary
A New Method for Stock Price Prediction Based on MRFs and SSVM .............................................................818
Lin Lai, Chang Li, and Wen Long
A Feasible Direction Method for Optimization Problem with Orthogonal Constraint in Feature Selection .......................................................................................................................824
Jianyu Miao, Yong Shi, and Lingfeng Niu
A Novel l0-Constrained Gaussian Graphical Model for Anomaly Localization ..................................................830
Dzung T. Phan, Tsuyoshi Idé, Jayant Kalagnanam, Matt Menickelly, and Katya Scheinberg
DMIP - Data Mining in Practice: Automation and Cost
An Empirical Evaluation of Techniques for Feature Selection with Cost ................................................................. 834
    Stephen Adams, Ryan Meekins, and Peter A. Beling
The Mean and Median Criteria for Kernel Bandwidth Selection for Support Vector Data Description .................................................. 842
    Arin Chaudhuri, Deovrat Kakde, Carol Sadek, Laura Gonzalez, and Seunghyun Kong
Combining Active Learning and Semi-Supervised Learning by Using Selective Label Spreading .................................................. 850
    Xu Chen and Tao Wang
Robust Self-Tuning Sparse Subspace Clustering ............................................................................................................. 858
    Guangtao Wang, Jiayu Zhou, Jingjie Ni, Tingjin Luo, Wei Long, Hai Zhen, Gao Cong, and Jieping Ye

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

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

BDDSR - Workshop on Big Data & Data Science in Retail
Mobile E-Commerce Data Processing Using Relational Memory ......................................................................................... 910
    Parham Aarabi
A Pattern Tree Based Method for Mining Conditional Contrast Patterns of Multi-source Data ........................................................................................................................................................................... 916
    Li Li, Sarah Erfani, and Christopher Leckie
Improving Multivariate Time Series Forecasting with Random Walks with Restarts on Causality Graphs ....................................................................................................................................................924
   Piotr Przymus, Youssef Hmamouche, Alain Casali, and Lotfi Lakhal

D3M - Data-driven Discovery of Models
Dataset Selection for Controlling Swarms by Visual Demonstration .........................................................................................................................932
   Karan Kumar Budhraja and Tim Oates
Discrimination at the Edge of Noise: A Hilbert Space of Stationary Ergodic Processes .................................................................942
   Ishanu Chattopadhyay
Ranking from Crowdsourced Pairwise Comparisons via Smoothed Matrix Manifold Optimization ........................................................................................................949
   Jialin Dong, Kai Yang, and Yuanming Shi
Feature Selection in Learning Using Privileged Information .................................................................................................................................957
   Rauf Izmailov, Blerta Lindqvist, and Peter Lin
A Central Limit Theorem for an Omnibus Embedding of Multiple Random Dot Product Graphs ..........964
   Keith Levin, Avanti Athreya, Minh Tang, Vince Lyzinski, and Carey E. Priebe

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

PAIS - Privacy and Anonymity in the Information Society
Factor Analysis for Anonymization .................................................................................................................................................................984
   Aida Calviño, Palmira Aldeguer, and Josep Domingo-Ferrer
Privacy-Preserving Big Data Stream Mining: Opportunities, Challenges, Directions ..................................................................................................992
   Alfredo Cuzzocrea
Steered Microaggregation: A Unified Primitive for Anonymization of Data Sets and Data Streams .........................................................................................................................995
   Josep Domingo-Ferrer and Jordi Soria-Comas
Personalized Anonymization for Set-Valued Data by Partial Suppression ...............................................................1003
  Takuma Nakagawa, Hiromi Arai, and Hiroshi Nakagawa

Revealing the Unseen: How to Expose Cloud Usage While Protecting User Privacy ..............................................1011
  Ata Turk, Mayank Varia, and Georgios Kellaris

**ACUMEN - Data Science for Human Performance in Social Networks**

Co-Training for Demographic Classification Using Deep Learning from Label Proportions .................................................................1017
  Ehsan Mohammady Ardehaly and Aron Culotta

EmTaggeR: A Word Embedding Based Novel Method for Hashtag Recommendation on Twitter .................................................................1025
  Kuntal Dey, Ritvik Shrivastava, Saroj Kaushik, and L. Venkata Subramaniam

Discovering Cooperative Structure Among Online Items for Attention Dynamics .................................................................1033
  Kanji Matsutani, Masahito Kumano, Masahiro Kimura, Kazumi Saito, Kouzou Ohara, and Hiroshi Motoda

Live on TV, Alive on Twitter: Quantifying Continuous Partial Attention of Viewers During Live Television Telecasts .................................................................1042
  Rohit Saxena, Savita Bhat, and Niranjan Pedanekar

Are Words Commensurate with Actions? Quantifying Commitment to a Cause from Online Public Messaging .................................................................1050
  Zhao Wang, Jennifer Cutler, and Aron Culotta

**DMCIS - Data Mining for Cyberphysical and Industrial Systems**

Anomaly Detection for a Water Treatment System Using Unsupervised Machine Learning .................................................................1058
  Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M. Poskitt, and Jun Sun

Pattern-Based Contextual Anomaly Detection in HVAC Systems .................................................................1066
  Mohsin Munir, Steffen Erkel, Andreas Dengel, and Sheraz Ahmed

An Adaptive Modeling Framework for Bivariate Data Streams with Applications to Change Detection in Cyber-Physical Systems .................................................................1074
  Joshua Plasse, Jordan Noble, and Kary Myers

Data-Driven Anomaly Detection for Power System Generation Control .................................................................1082
  Pengyuan Wang, Manimaran Govindarasu, Aditya Ashok, Siddharth Sridhar, and David McKinnon

**PhD Forum**

Dependency Anomaly Detection for Heterogeneous Time Series: A Granger-Lasso Approach .................................................................1090
  Sahar Behzadi, Katerina Hlaváčková-Schindler, and Claudia Plant

Deep Learning Solutions to Computational Phenotyping in Health Care .................................................................1100
  Zhengping Che and Yan Liu
Aggregation and Disaggregation of Information: A Holistic View ........................................................................................................1110
Yuyue Chen and Chuanren Liu

Deep Learning for Pulmonary Nodule CT Image Retrieval — An Online Assistance
System for Novice Radiologists .................................................................................................................................1112
Daniel Perez Ibanez, Jiang Li, Yuzhong Shen, Joan Dayanghirang, Shengli Wang, and Zezhong Zheng

Development of an Interpretable Neural Network Model for Creation of Polarity Concept
Dictionaries ................................................................................................................................................................1122
Tomoki Ito, Hiroki Sakaji, Kiyoshi Izumi, Kota Tsubouchi, and Tatsuo Yamashita

Uncovering Teamwork in Networks — Prediction, Optimization and Explanation ....................................................1 132
Liangyue Li and Hanghang Tong

A Multilevel NER Framework for Automatic Clinical Name Entity Recognition ..................................................1134
Thoai Man Luu, Robert Phan, Rachel Davey, and Girija Chetty

Network Embedding with Centrality Information ........................................................................................................1144
Yao Ma, Suhang Wang, and Jiliang Tang

Predicting Weight Loss with Enemble Methods ........................................................................................................1146
Zhiwei Wang and Jiliang Tang

Crowdsourcing Data Science for Innovation .............................................................................................................1148
Wangcheng Yan, Wenjun Zhou, Paolo Letizia, and Bogdan Bichescu

Exploring Transfer Learning for Crime Prediction .....................................................................................................1158
Xiangyu Zhao and Jiliang Tang

Demo Session

FRS: Fast Range Search by Pruning Unnecessary Distance Computations Based on K-D Tree .................................................................1160
Yewang Chen, Jai Puneet Singh, Lida Zhou, and Nizar Bouguila

Pi-CEP: Predictive Complex Event Processing Using Range Queries over Historical Pattern Space .....................................................................1166
Syed Gillani, Abderrahmen Kammoun, Kamal Singh, Julien Subercaze, Christophe Gravier, Jacques Fayolle, and Frédérique Laforest

Mining Human Mobility to Quantify Performance Status ...........................................................................................1172
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

TweeLoc: A System for Geolocalizing Tweets at Fine-Grain ....................................................................................1178
Pavlos Paraskevopoulos, Giovanni Pellegrini, and Themis Palpanas

Additional Paper

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

Author Index