

2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW 2016)

**Barcelona, Spain
12-15 December 2016**

Pages 1-670



**IEEE Catalog Number: CFP1656B-POD
ISBN: 978-1-5090-5911-9**

**Copyright © 2016 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:	CFP1656B-POD
ISBN (Print-On-Demand):	978-1-5090-5911-9
ISBN (Online):	978-1-5090-5910-2
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

2016 IEEE 16th International Conference on Data Mining Workshops

ICDMW 2016

Table of Contents

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

CLOUDBMINE + DMIOt

CLOUDBMINE — Data Mining Systems and their Applications on the Cloud

Database Integrated Analytics Using R: Initial Experiences with SQL-Server + R	1
<i>Josep Ll. Berral and Nicolas Poggi</i>	
Parallelized Frequent Item Set Mining Using a Tall and Skinny Matrix	8
<i>D. Pooja Janakiram</i>	
A Probabilistic View of Neighborhood-Based Recommendation Methods	14
<i>Jun Wang and Qiang Tang</i>	
Discovering Multi-type Correlated Events with Time Series for Exception Detection of Complex Systems	21
<i>Peng Xun, Pei-Dong Zhu, Cun-Lu Li, and Hao-Yang Zhu</i>	

DMIOt — Data Mining for Internet of Things

Detecting Performance Degradation of Software-Intensive Systems in the Presence of Trends and Long-Range Dependence	29
<i>Alexey Artemov and Evgeny Burnaev</i>	
Scalable Online-Offline Stream Clustering in Apache Spark	37
<i>Omar Backhoff and Eirini Ntoutsi</i>	
Distributed Mining and Modeling of Dynamic Lead-Lag Relations in Evolving Entities	45
<i>Tian Guo, Jean-Paul Calbimonte, and Karl Aberer</i>	

Event Detection for Urban Dynamic Data Streams	53
<i>Septimiu Nechifor, Ioana Stefan, Marten Fischer, and Dan Puiu</i>	

DaMNet — Data Mining in Networks

Segmenting Sequences of Node-Labeled Graphs	61
<i>Sorour E. Amiri, Liangzhe Chen, and B. Aditya Prakash</i>	
Inference of Partial Canonical Correlation Networks with Application to Stock Market Portfolio Selection	69
<i>Gregory Breard and Natallia Katenka</i>	
Overlapping Community Detection by Local Decentralised Vertex-Centred Process	77
<i>Maël Canu, Marie-Jeanne Lesot, and Adrien Revault D’Allonnes</i>	
Query-Based Evolutionary Graph Cuboid Outlier Detection	85
<i>Ayushi Dalmia, Manish Gupta, and Vasudeva Varma</i>	
Vertex Centric Asynchronous Belief Propagation Algorithm for Large-Scale Graphs	93
<i>Gabriel P. Gimenes, Hugo Gualdron, and Jose F. Rodrigues-Jr</i>	
Text Network Exploration via Heterogeneous Web of Topics	99
<i>Junxian He, Ying Huang, Changfeng Liu, Jiaming Shen, Yuting Jia, and Xinbing Wang</i>	
Classification of Normal and Pathological Brain Networks Based on Similarity in Graph Partitions	107
<i>Anvar Kurmukov, Yulia Dodonova, and Leonid Zhukov</i>	
Finding Heaviest k-Subgraphs and Events in Social Media	113
<i>Matthaios Letsios, Oana Denisa Balalau, Maximilien Danisch, Emmanuel Orsini, and Mauro Sozio</i>	
Knowledge Graph Constraints for Multi-label Graph Classification	121
<i>Martin Ringsquandl, Steffen Lamparter, Ingo Thon, Raffaello Lepratti, and Peer Kröger</i>	
Predicting User Roles in Social Networks Using Transfer Learning with Feature Transformation	128
<i>Jun Sun, Jérôme Kunegis, and Steffen Staab</i>	

DAPS — Data Mining for the Analysis of Performance and Success

From Training to Match Performance: A Predictive and Explanatory Study on Novel Tracking Data	136
<i>Javier Fernández, Daniel Medina, Antonio Gómez, Marta Arias, and Ricard Gavaldá</i>	
Competition-Wide Evaluation of Individual and Team Movements in Soccer	144
<i>László Gyarmati and Mohamed Hefeeda</i>	
Football Market Strategies: Think Locally, Trade Globally	152
<i>Giulio Rossetti and Vincenzo Caproni</i>	

Characterization of In-season Elite Football Trainings by GPS Features: The Identity Card of a Short-Term Football Training Cycle	160
<i>Alessio Rossi, Enrico Perri, Athos Trecroci, Marco Savino, Giampietro Alberti, and Marcello Fedon Iaia</i>	
Towards Data-Driven Football Player Assessment	167
<i>Rade Stanojevic and László Gyarmati</i>	
Stick Must Fall: Using Machine Learning to Predict Human Error in Virtual Balancing Task	173
<i>Irina Zgonnikova, Arkady Zgonnikov, and Shigeru Kanemoto</i>	

DINA — Data Integration and Applications

Towards Information Profiling: Data Lake Content Metadata Management	178
<i>Ayman Alserafi, Alberto Abelló, Oscar Romero, and Toon Calders</i>	
Historical Data Integration a Study of WWI Canadian Soldiers	186
<i>Luiza Antonie, Harshavardhan Gadgil, Gary Grewal, and Kris Inwood</i>	
Holistic Entity Clustering for Linked Data	194
<i>Markus Nentwig, Anika Groß, and Erhard Rahm</i>	
Cluster-Based Sorted Neighborhood for Efficient Duplicate Detection	202
<i>Ahmad Samiei and Felix Naumann</i>	
Integrated Analysis of Research Publications and Patents for Strategic Decision Making	210
<i>Nidhi Saraswat, Lipika Dey, Ishan Verma, and Hemant Gupta</i>	
Randomized Response and Balanced Bloom Filters for Privacy Preserving Record Linkage	218
<i>Rainer Schnell and Christian Borgs</i>	
A Probabilistic Address Parser Using Conditional Random Fields and Stochastic Regular Grammar	225
<i>Minlue Wang, Valeriia Haberland, Amos Yeo, Andrew Martin, John Howroyd, and J. Mark Bishop</i>	

DMBIH — Data Mining in Biomedical Informatics and Healthcare

Construction of a Semi-Naive Model to Predict Early Readmission of COPD Patients by Using Quality Care Information	233
<i>Pablo Bermejo, José A. Gámez, José M. Puerta, Marco A. Esquivias, and Pedro J. Tárrega</i>	
Medical Image Denoising Using Convolutional Denoising Autoencoders	241
<i>Lovedeep Gondara</i>	

Machine Learning for Modeling the Biomechanical Behavior of Human Soft Tissue	247
<i>José D. Martín-Guerrero, María J. Rupérez-Moreno, Francisco Martínez-Martínez, Delia Lorente-Garrido, Antonio J. Serrano-López, Carlos Monserrat, Sandra Martínez-Sanchis, and Marcelino Martínez-Sober</i>	
Applying Conditional Independence Maps to Improve Sepsis Prognosis	254
<i>Carles Morales, Vicent Ribas, and Alfredo Vellido</i>	
Ligand-Based Virtual Screening with Co-regularised Support Vector Regression	261
<i>Katrin Ullrich, Michael Kamp, Thomas Gärtner, Martin Vogt, and Stefan Wrobel</i>	
DMCS — Data Mining for Cyber Security	
Uncertainty Aware Clustering for Behaviour in Enterprise Networks	269
<i>Maha Bakoben, Niall Adams, and Anthony Bellotti</i>	
One-Class SVM with Privileged Information and Its Application to Malware Detection	273
<i>Evgeny Burnaev and Dmitry Smolyakov</i>	
Adaptive Threshold Selection for Trust-Based Detection Systems	281
<i>Younghun Chae, Natallia Katenka, and Lisa Dipippo</i>	
The Time-Varying Dependency Patterns of NetFlow Statistics	288
<i>Alexander Gibberd, Marina Evangelou, and James D. B. Nelson</i>	
Learning a Secure Classifier against Evasion Attack	295
<i>Zeinab Khorshidpour, Sattar Hashemi, and Ali Hamzeh</i>	
Detection of Cyber-Physical Faults and Intrusions from Physical Correlations	303
<i>Andrey Y. Lokhov, Nathan Lemons, Thomas C. McAndrew, Aric Hagberg, and Scott Backhaus</i>	
Correlation-Based Streaming Anomaly Detection in Cyber-Security	311
<i>Jordan Noble and Niall M. Adams</i>	
Smart Phone User Behaviour Characterization Based on Autoencoders and Self Organizing Maps	319
<i>Deepthi Rajashekar, A. Nur Zincir-Heywood, and Malcolm I. Heywood</i>	
EdgeCentric: Anomaly Detection in Edge-Attributed Networks	327
<i>Neil Shah, Alex Beutel, Bryan Hooi, Leman Akoglu, Stephan Gunnemann, Disha Makhija, Mohit Kumar, and Christos Faloutsos</i>	
Authenticating Web User’s Identity through Browsing Sequences Modeling	335
<i>Peihai Zhao, Chungang Yan, and Changjun Jiang</i>	

DMED I — Data Mining in Emerging Domains I

DWA — Data Wrangling Automation

Using Machine Learning to Accelerate Data Wrangling	343
<i>Shilpi Ahuja, Mary Roth, Rashmi Gangadharaiah, Peter Schwarz, and Rafael Bastidas</i>	
Mining the Dark Web: Drugs and Fake Ids	350
<i>Andres Baravalle, Mauro Sanchez Lopez, and Sin Wee Lee</i>	
Toward Representation Independent Analytics over Structured Data	357
<i>Yodsawalai Chodpathumwan, Jose Picado, Arash Termehchy, Alan Fern, and Yizhou Sun</i>	

MDDM — Maritime Domain Data Mining

Mining Vessel Tracking Data for Maritime Domain Applications	361
<i>Alfredo Alessandrini, Marlene Alvarez, Harm Greidanus, Vincenzo Gammieri, Virginia Fernandez Arguedas, Fabio Mazzarella, Carlos Santamaria, Mattia Stasolla, Dario Tarchi, and Michele Vespe</i>	
On the Interest of Data Mining for an Integrity Assessment of AIS Messages	368
<i>Clément Iphar, Aldo Napoli, and Cyril Ray</i>	
Scalable and Distributed Sea Port Operational Areas Estimation from AIS Data	374
<i>Leonardo M. Millefiori, Dimitrios Zissis, Luca Cazzanti, and Gianfranco Arcieri</i>	

DMED II — Data Mining in Emerging Domains II

DaMEMO — Data Mining for Energy Modelling and Optimization

Application of Fuzzy C-Means for Proactive Clustering of Electrical Power Systems	382
<i>M. Abdallatif, S. Schramm, and J. Götze</i>	
Online Outlier Detection of Energy Data Streams Using Incremental and Kernel PCA Algorithms	390
<i>Jeremiah D. Deng</i>	
Adaptive Time Series Forecasting of Energy Consumption Using Optimized Cluster Analysis	398
<i>Peter Laurinec, Marek Lóderer, Petra Vrablecová, Mária Lucká, Viera Rozinajová, and Anna Bou Ezzeddine</i>	
Structural Combination of Neural Network Models	406
<i>Juan Rendon and Lilian M. de Menezes</i>	
Incremental Time Series Prediction Using Error-Driven Informed Adaptation	414
<i>Petra Vrablecová, Viera Rozinajová, and Anna Bou Ezzeddine</i>	

DMiP — Data Mining in Politics

Analysing Political Opinions Using Redescription Mining	422
<i>Esther Galbrun and Pauli Miettinen</i>	
Leave or Remain? Deciphering Brexit Deliberations on Twitter	428
<i>Apalak Khatua and Aparup Khatua</i>	
Change-Point Analysis of the Public Mood in UK Twitter during the Brexit Referendum	434
<i>Thomas Lansdall-Welfare, Fabon Dzogang, and Nello Cristianini</i>	
Structural Patterns in the Rise of Germany’s New Right on Facebook	440
<i>Sebastian Schelter, Felix Biessmann, Malisa Zobel, and Nedelina Teneva</i>	

DMHAA — Data Mining in Human Activity Analysis

Weakly Supervised Hand Pose Recovery with Domain Adaptation by Low-Rank Alignment	446
<i>Chaoqun Hong, Jun Yu, Rongsheng Xie, and Dapeng Tao</i>	
Link Prediction in the Twitter Mention Network: Impacts of Local Structure and Similarity of Interest	454
<i>Hadrien Hours, Eric Fleury, and Márton Karsai</i>	
Gender Classification by Deep Learning on Millions of Weakly Labelled Images	462
<i>Sen Jia, Thomas Lansdall-Welfare, and Nello Cristianini</i>	
Real-Time Top-View People Counting Based on a Kinect and NVIDIA Jetson TK1 Integrated Platform	468
<i>Guangqin Li, Peng Ren, Xinrong Lyu, and He Zhang</i>	
Online Semi-Supervised Multi-task Distance Metric Learning	474
<i>Ya Li and Dacheng Tao</i>	
Local Structure Preserving Using Manifold Regularization and Pairwise Constraints for Action Recognition	480
<i>Xueqi Ma, Zhengyang Cao, and Weifeng Liu</i>	
Stereotypical Motor Movement Detection in Dynamic Feature Space	487
<i>Nastaran Mohammadian Rad, Seyed Mostafa Kia, Calogero Zarbo, Giuseppe Jurman, Paola Venuti, and Cesare Furlanello</i>	
Robust Automated Human Activity Recognition and Its Application to Sleep Research	495
<i>Aarti Sathyanarayana, Ferda Ofli, Luis Fernandez-Luque, Jaideep Srivastava, Ahmed Elmagarmid, Teresa Arora, and Shahrads Taheri</i>	
Sequence-to-Sequence Model with Attention for Time Series Classification	503
<i>Yujin Tang, Jianfeng Xu, Kazunori Matsumoto, and Chihiro Ono</i>	

DMS — Data Mining for Service

Risk-Aware Dynamic Reserve Prices of Programmatic Guarantee in Display Advertising	511
<i>Bowei Chen</i>	
A Novel Method for Assessing Event Impacts on Event-Driven Time Series	519
<i>Lianhua Chi, Saket Sathe, Bo Han, and Yun Wang</i>	
Temporal Data Mining on the Stay Time of Outpatients and Treatment Processes	527
<i>Shoji Hirano</i>	
A Deep Learning Approach for the Prediction of Retail Store Sales	531
<i>Yuta Kaneko and Katsutoshi Yada</i>	
Website Classification Using Latent Dirichlet Allocation and Its Application for Internet Advertising	538
<i>Sotaro Katsumata, Eiji Motohashi, Akihiro Nishimoto, and Eiji Toyosawa</i>	
Personal Identification with Face and Voice Features Extracted through Kinect Sensor	545
<i>Eisuke Kita, Yi Zuo, Fumiya Saito, and Xuanang Feng</i>	
Data Mining Based Product Marketing Technique for Banking Products	552
<i>Merve Mitik, Ozan Korkmaz, Pinar Karagoz, Ismail Hakki Toroslu, and Ferhat Yucel</i>	
Interactive Independent Topic Analysis for Service	560
<i>Takahiro Nishigaki, Katsumi Nitta, and Takashi Onoda</i>	
Activity Detection from Email Meta-Data Clustering	568
<i>Mayur Patidar, Shaurya Rohatgi, Ashish Chaudhary, Mahesh P. Singh, Puneet Agarwal, and Gautam Shroff</i>	
Feature Selection in Environments with Limited Voluntary Information Sharing	576
<i>Nicholay Topin, Karan K. Budhraj, and Tim Oates</i>	
Domain Driven Forecasting: Applying Theory Into Practice	584
<i>Praveen Venkateswaran, Vikrant Shimpi, Maitreya Natu, and Vaishali Sadaphal</i>	
Multiscale Analysis of Supply Network at Central Region in Japan	592
<i>Yi Zuo and Yuya Kajikawa</i>	

DSBDA — Data Science and Big Data Analytics

Visual Context Learning with Big Data Analytics	600
<i>Mayanka Chandrashekar and Yugyung Lee</i>	
Finding Large Matchings in Semi-Streaming	608
<i>Hossein Esfandiari, MohammadTaghi Hajiaghayi, and Morteza Monemizadeh</i>	
Effective and Unsupervised Fractal-Based Feature Selection for Very Large Datasets: Removing Linear and Non-linear Attribute Correlations	615
<i>Antonio C. Fraideinberze, José F. Rodrigues Jr., and Robson L. F. Cordeiro</i>	

Tovel: Distributed Graph Clustering for Word Sense Disambiguation	623
<i>Alessio Guerrieri, Fatemeh Rahimian, Sarunas Girdzijauskas, and Alberto Montresor</i>	
Will I Get in? Modeling the Graduate Admission Process for American Universities	631
<i>Narender Gupta, Aman Sawhney, and Dan Roth</i>	
Selecting the Top-k Discriminative Features Using Principal Component Analysis	639
<i>Aminata Kane and Nematollaah Shiri</i>	
Score Look-Alike Audiences	647
<i>Qiang Ma, Eeshan Wagh, Jiayi Wen, Zhen Xia, Robert Ormandi, and Datong Chen</i>	
Designing Sketches for Similarity Filtering	655
<i>Vladimir Mic, David Novak, and Pavel Zezula</i>	
Tracking Language Mobility in the Twitter Landscape	663
<i>Izabela Moise, Edward Gaere, Ruben Merz, Stefan Koch, and Evangelos Pournaras</i>	
Relative Label Encoding for the Prediction of Airline Passenger Nationality	671
<i>Alejandro Mottini and Rodrigo Acuna-Agost</i>	
WEFEST: Word Embedding Feature Extension for Short Text Classification	677
<i>Lei Sang, Fei Xie, Xiaojian Liu, and Xindong Wu</i>	
Knowledge Representation Learning via Dynamic Relation Spaces	684
<i>Zhen Tan, Xiang Zhao, Yang Fang, Weidong Xiao, and Jiuyang Tang</i>	
Steady Patterns	692
<i>Willy Ugarte, Alexandre Termier, and Miguel Santana</i>	
Mining Effective Subsequences with Application in Marketing Attribution	700
<i>Zi Yin, Ying Li, Pietro Mazzoleni, and Yuanyuan Shen</i>	
HDM — High Dimensional Data Mining	
Random Projection Clustering on Streaming Data	708
<i>Lee A. Carraher, Philip A. Wilsey, Anindya Moitra, and Sayantan Dey</i>	
An Empirical Analysis of Hubness in Unsupervised Distance-Based Outlier Detection	716
<i>Arthur Flexer</i>	
Topic Extraction Method from Millions of Tweets Based on Fast Feature Selection Technique CWC	724
<i>Takako Hashimoto, Dave Shepard, Tetsuji Kuboyama, and Kilho Shin</i>	
Regression on High-Dimensional Inputs	732
<i>Alexander Kuleshov and Alexander Bernstein</i>	
Clustering Based on MultiView Diffusion Maps	740
<i>Ofir Lindenbaum, Arie Yeredor, and Amir Averbuch</i>	
Aggregating Tree for Searching in Billion Scale High Dimensional Data	748
<i>Shicong Liu, Junru Shao, and Hongtao Lu</i>	

Semi-Supervised Similarity Preserving Co-Selection	756
<i>Raywat Makkhongkaew and Khalid Benabdeslem</i>	
Robust Local Scaling Using Conditional Quantiles of Graph Similarities	762
<i>Jayaraman J. Thiagarajan, Prasanna Sattigeri, Karthikeyan Natesan Ramamurthy, and Bhavya Kailkhura</i>	
Modelling Sentence Generation from Sum of Word Embedding Vectors as a Mixed Integer Programming Problem	770
<i>Lyndon White, Roberto Togneri, Wei Liu, and Mohammed Bennamoun</i>	

MoDAT — Data Market for Co-Evolution of Sciences and Business

Method for Extraction of Purchase Behavior and Product Character Using Dynamic Topic Model	778
<i>Mamoru Emoto</i>	
Inferring Variable Labels Considering Co-occurrence of Variable Labels in Data Jackets	783
<i>Teruaki Hayashi</i>	
The Prominent Role of Probe Data and Big Data in Modern Technology	788
<i>Hideki Isetani</i>	
ID-Link, an Enabler for Medical Data Marketplace	792
<i>Ryuji Ito</i>	
Predicting the Outcome of Startups: Less Failure, More Success	798
<i>Amar Krishna, Ankit Agrawal, and Alok Choudhary</i>	
Aware Environment for Workshop with Game Storming	806
<i>Noriyuki Kushiro, Tatsuya Ehira, and Rei Kaihara</i>	
BRPS: A Big Data Placement Strategy for Data Intensive Applications	813
<i>Lihui Liu, Junping Song, Haibo Wang, and Pin Lv</i>	
Data Exchange Platform to Fight Insurance Fraud on Blockchain	821
<i>Indranil Nath</i>	
Visualization Method Based on Contour Map	826
<i>Rui Yang and Yukio Ohsawa</i>	

OEDM — Optimization Based Techniques for Emerging Data Mining

MoCham: Robust Hierarchical Clustering Based on Multi-objective Optimization	831
<i>Tomas Barton, Tomas Bruna, and Pavel Kordik</i>	
Structural Nonparallel Support Vector Machine Based on LSH for Large-Scale Prediction	839
<i>Dandan Chen</i>	
Laplacian SVM for Learning from Label Proportions	847
<i>Limeng Cui, Zhensong Chen, Fan Meng, and Yong Shi</i>	

Incorporate Hashing with Multi-view Learning	853
<i>Jingjing Tang and Dewei Li</i>	

PDDM — Privacy and Discrimination in Data Mining

Model-Based Approaches for Independence-Enhanced Recommendation	860
<i>Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Issei Sato</i>	
Classification Rule Mining Supported by Ontology for Discrimination Discovery	868
<i>Binh Thanh Luong, Salvatore Ruggieri, and Franco Turini</i>	
Faster Univariate Microaggregation for Power Law Distributions: k-Degree-Anonymity for Big Graphs	876
<i>Julian Salas</i>	
Scalable Privacy-Preserving Linking of Multiple Databases Using Counting Bloom Filters	882
<i>Dinusha Vatsalan, Peter Christen, and Erhard Rahm</i>	

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

Research on Sentiment Analysis: The First Decade	890
<i>Oskar Ahlgren</i>	
Mining the Opinionated Web: Classification and Detection of Aspect Contexts for Aspect Based Sentiment Analysis	900
<i>Oscar Araque, Ganggao Zhu, Manuel García-Amado, and Carlos A. Iglesias</i>	
What Drives Consumer Choices? Mining Aspects and Opinions on Large Scale Review Data Using Distributed Representation of Words	908
<i>Kasturi Bhattacharjee and Linda Petzold</i>	
Bayesian Deep Convolution Belief Networks for Subjectivity Detection	916
<i>Iti Chaturvedi, Erik Cambria, Soujanya Poria, and Rajiv Bajpai</i>	
Sentiment Lexica from Paired Comparisons	924
<i>Christoph Dalitz and Katrin E. Bednarek</i>	
Seasonal Fluctuations in Collective Mood Revealed by Wikipedia Searches and Twitter Posts	931
<i>Fabon Dzogang, Thomas Lansdall-Welfare, and Nello Cristianini</i>	
The Truth and Nothing But the Truth: Multimodal Analysis for Deception Detection	938
<i>Mimansa Jaiswal, Sairam Tabibu, and Rajiv Bajpai</i>	
Scalable and Real-Time Sentiment Analysis of Twitter Data	944
<i>Maria Karanasou, Anneta Ampla, Christos Doulkeridis, and Maria Halkidi</i>	
Multi-sentiment Modeling with Scalable Systematic Labeled Data Generation via Word2Vec Clustering	952
<i>Dhruv Mayank, Kanchana Padmanabhan, and Koushik Pal</i>	

SmartVideoRanking: Video Search by Mining Emotions from Time-Synchronized Comments	960
<i>Kosetsu Tsukuda, Hamasaki Masahiro, and Masataka Goto</i>	
Modeling Satire in English Text for Automatic Detection	970
<i>Aishwarya N. Reganti, Tushar Maheshwari, Upendra Kumar, Amitava Das, and Rajiv Bajpai</i>	
Lexicon Knowledge Extraction with Sentiment Polarity Computation	978
<i>Zhaoxia Wang, Victor Joo Chuan Tong, Pingcheng Ruan, and Fang Li</i>	

SERecSys — Semantics-Enabled Recommender Systems

Exploiting a Determinant-Based Metric to Evaluate a Word-Embeddings Matrix of Items	984
<i>Ludovico Boratto, Salvatore Carta, Gianni Fenu, and Roberto Saia</i>	
Semantic Enabled Recommender System for Micro-Blog Users	992
<i>Stefano Faralli, Giorgia Di Tommaso, and Paola Velardi</i>	
SemStim: Exploiting Knowledge Graphs for Cross-Domain Recommendation	999
<i>Benjamin Heitmann and Conor Hayes</i>	
Context-Specific Recommendation System for Predicting Similar PubMed Articles	1007
<i>Shahin Mohammadi, Sudhir Kylasa, Giorgos Kollias, and Ananth Grama</i>	
SeBPR: Semantics Enhanced Bayesian Personalized Ranking with Comparable Item Pairs	1015
<i>Penghua Yu and Lanfen Lin</i>	

SOMERIS — Social Media and Risk

Learning from User Workflows for the Characterization and Prediction of Software Crashes	1023
<i>Chloé Adam, Antoine Aliotti, and Paul-Henry Cournède</i>	
Microblog Sentiment Topic Model	1031
<i>Aman Ahuja, Wei Wei, and Kathleen M. Carley</i>	
Graph-Based Term Weighting Scheme for Topic Modeling	1039
<i>Giannis Bekoulis and François Rousseau</i>	
Ensemble of Heterogeneous Classifiers for Improving Automated Tweet Classification	1045
<i>Renhao Cui, Gagan Agrawal, Rajiv Ramnath, and Vinh Khuc</i>	
Identifying Warning Behaviors of Violent Lone Offenders in Written Communication	1053
<i>Lisa Kaati, Amendra Shrestha, and Tony Sardella</i>	
Investors Attention and the Effects on Stock Market: An Empirical Study Based on Stock Forum	1061
<i>Wen Long, Lijing Guan, and Lingxiao Cui</i>	

Time-Based Ensembles for Prediction of Rare Events in News Stream	1066
<i>Nuno Moniz, Luis Torgo, and Magdalini Eirinaki</i>	
Fraud Detection in Voice-Based Identity Authentication Applications and Services	1074
<i>Saeid Safavi, Hock Gan, Iosif Mporas, and Reza Sotudeh</i>	
SSTDM — Spatial and Spatiotemporal Data Mining	
Discovering Spatial Regions of High Correlation	1082
<i>Prerna Agarwal, Richa Verma, and Venkata M. V. Gunturi</i>	
A Graph-Based Approach to Spatiotemporal Event Sequence Mining	1090
<i>Berkay Aydin and Rafal A. Angryk</i>	
Hierarchical Aggregation Approach for Distributed Clustering of Spatial Datasets	1098
<i>Malika Bendecheche, Nhien-An Le-Khac, and M-Tahar Kechadi</i>	
Markov Switching Copula Models for Longitudinal Data	1104
<i>Alfredo Cuesta-Infante and Kalyan Veeramachaneni</i>	
Location Recommendations for New Businesses Using Check-in Data	1110
<i>Bahaeddin Eravci, Neslihan Bulut, Cagri Etemoglu, and Hakan Ferhatosmanoğlu</i>	
Generating Time-Varying Road Network Data Using Sparse Trajectories	1118
<i>Elif Eser, Furkan Kocayusufoğlu, Bahaeddin Eravci, Hakan Ferhatosmanoğlu, and Josep L. Larriba-Pey</i>	
A Pattern Growth-Based Approach for Mining Spatiotemporal Co-occurrence Patterns	1125
<i>Shah Muhammad Hamdi, Berkay Aydin, and Rafal A. Angryk</i>	
Correlating Social Interconnections of Users with Spatio-Temporal Check-Ins Behavior	1133
<i>Sonia Khetarpaul, S. K. Gupta, and L. Venkata Subramaniam</i>	
Flexible Similarity Search for Enriched Trajectories	1139
<i>Hideaki Ohashi, Toshiyuki Shimizu, and Masatoshi Yoshikawa</i>	
TrendTracker: Modelling the Motion of Trends in Space and Time	1145
<i>Klaus Arthur Schmid, Christian Frey, Fengchao Peng, Michael Weiler, Andreas Zufle, Lei Chen, and Matthias Renz</i>	
A Genetic Algorithm to Discover Flexible Motifs with Support	1153
<i>Joan Serrà, Aleksandar Matic, Josep Lluís Arcos, and Alexandros Karatzoglou</i>	
Vehicle Classification from Low Frequency GPS Data	1159
<i>Matteo Simoncini, Francesco Sambo, Leonardo Taccari, Luca Bravi, Samuele Salti, and Alessandro Lori</i>	
Hierarchical Temporal Memory Method for Time-Series-Based Anomaly Detection	1167
<i>Jia Wu, Weiru Zeng, Zhe Chen, and Xue-Fei Tang</i>	
Spatio-Temporal Site Recommendation	1173
<i>Guolei Yang and Andreas Züfle</i>	

PhD Forum

Sacrificing Overall Classification Quality to Improve Classification Accuracy of Well-Sought Classes	1179
<i>Kevin M. Amaral, Ping Chen, Wei Ding, and Rajani Sadasivam</i>	
Towards Visualizing Hidden Structures	1183
<i>Remy Dautriche, Alexandre Termier, Renaud Blanch, and Miguel Santana</i>	
Regression Techniques for Modelling Conception in Seasonally Calving Dairy Cows	1191
<i>Caroline Fenlon, Luke O’Grady, Michael Doherty, Stephen Butler, Laurence Shalloo, and John Dunnion</i>	
A Semi-Supervised Ensemble Approach for Multi-label Learning	1197
<i>Ouadie Gharroudi, Haytham Elghazel, and Alexandre Aussem</i>	
A Novel Bayesian Ensemble Pruning Method	1205
<i>Zhengshen Jiang, Hongzhi Liu, Bin Fu, and Zhonghai Wu</i>	
Improving the Prediction Cost of Drift Handling Algorithms by Abstaining	1213
<i>Pierre-Xavier Loeffel, Vincent Lemaire, Christophe Marsala, and Marcin Detyniecki</i>	
Infer Mobility Patterns and Social Dynamics for Modelling Human Behaviour	1223
<i>Luca Luceri</i>	
Subspace Clustering Ensembles through Tensor Decomposition	1225
<i>Dominik Mautz, Christian Böhm, and Claudia Plant</i>	
Applying Deep Learning to Stereotypical Motor Movement Detection in Autism Spectrum Disorders	1235
<i>Nastaran Mohammadian Rad and Cesare Furlanello</i>	
Similarity Tree Pruning: A Novel Dynamic Ensemble Selection Approach	1243
<i>Anil Narassiguin, Haytham Elghazel, and Alex Aussem</i>	
Clustering with the Levy Walk: “Hunting” for Clusters	1251
<i>Benjamin Schelling and Claudia Plant</i>	
Centrality-Based Approach for Supervised Term Weighting	1261
<i>Niloofar Shanavas, Hui Wang, Zhiwei Lin, and Glenn Hawe</i>	
A Novel Pre-Classification Based kNN Algorithm	1269
<i>Huahua Xie, Dong Liang, Zhaojing Zhang, Hao Jin, Chen Lu, and Yi Lin</i>	

Demos

A Formation Energy Predictor for Crystalline Materials Using Ensemble Data Mining	1276
<i>Ankit Agrawal, Bryce Meredig, Chris Wolverton, and Alok Choudhary</i>	
Five Year Life Expectancy Calculator for Older Adults	1280
<i>Ankit Agrawal, Jason Mathias, David Baker, and Alok Choudhary</i>	

Financial Data Analysis with PGMs Using AMIDST	1284
<i>Rafael Cabañas, Ana M. Martínez, Andrés R. Masegosa, Darío Ramos-López, Antonio Samerón, Thomas D. Nielsen, Helge Langseth, and Anders L. Madsen</i>	
TENET: A Machine Learning-Based System for Target Characterization in Signaling Networks	1288
<i>Huey Eng Chua, Sourav S. Bhowmick, Lisa Tucker-Kellogg, and C. Forbes Dewey Jr.</i>	
PaTSI: Pattern Mining of Time Series of Satellite Images in Knime	1292
<i>Maxime Collin, Frédéric Flouvat, and Nazha Selmaoui-Folcher</i>	
airVLC: An Application for Visualizing Wind-Sensitive Interpolation of Urban Air Pollution Forecasts	1296
<i>Lidia Contreras-Ochando and Cèsar Ferri</i>	
Explode: An Extensible Platform for Differentially Private Data Analysis	1300
<i>Emir Esmerdag, Mehmet Emre Gursoy, Ali Inan, and Yucel Saygin</i>	
Cognito: Automated Feature Engineering for Supervised Learning	1304
<i>Udayan Khurana, Deepak Turaga, Horst Samulowitz, and Srinivasan Parthasarathy</i>	
Risk Mining: Company-Risk Identification from Unstructured Sources	1308
<i>Timothy Nugent and Jochen L. Leidner</i>	
MobInsight: Understanding Urban Mobility with Crowd-Powered Neighborhood Characterizations	1312
<i>Souneil Park, Marc Bourqui, and Enrique Frias-Martinez</i>	
Graph Mining for Complex Data Analytics	1316
<i>André Petermann, Martin Junghanns, Stephan Kemper, Kevin Gómez, Niklas Teichmann, and Erhard Rahm</i>	
FacetsViewer: A Tool for Multi-faceted Decomposition of Complex Networks	1320
<i>Boon-Siew Seah, Sourav S. Bhowmick, Huey Eng Chua, Mengxuan Chen, and C. Forbes Dewey Jr.</i>	
SpreadViz: Analytics and Visualization of Spreading Processes in Social Networks	1324
<i>Konstantinos Skianis, Maria-Evgenia G. Rossi, Fragkiskos D. Malliaros, and Michalis Vazirgiannis</i>	
DFA-G: A Unified Programming Model for Vertex-Centric Parallel Graph Processing	1328
<i>Bo Suo, Jing Su, Qun Chen, Zhanhuai Li, and Wei Pan</i>	
Informed Design Platform: Interpreting “Big Data” to Adaptive Place Designs	1332
<i>Linlin You and Bige Tunçer</i>	

Author Index