

# **2015 IEEE International Conference on Data Mining Workshop (ICDMW 2015)**

**Atlantic City, New Jersey, USA  
14-17 November 2015**

**Pages 1-854**



**IEEE Catalog Number: CFP1556B-POD  
ISBN: 978-1-4673-8494-0**

**Copyright © 2015 by the Institute of Electrical and Electronic 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 publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP1556B-POD
ISBN (Print-On-Demand):	978-1-4673-8494-0
ISBN (Online):	978-1-4673-8493-3
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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2015 IEEE 15th International Conference on Data Mining Workshops

## ICDMW 2015

### Table of Contents

Message from the Conference Chairs .....	xxi
Message from the Workshop Co-Chairs .....	xxiii
Organization .....	xxv

---

#### Analyzing Events Using Social Media Data

Automatic Visual Analysis of Real-World Events Covered By Social Media Using Convolutional Neural Networks .....	1
<i>Henning Hamer, Andreas Merentitis, Nikolaos Frangiadakis, and Sergey Shukanov</i>	
Event Detection from Millions of Tweets Related to the Great East Japan Earthquake Using Feature Selection Technique .....	7
<i>Takako Hashimoto, Dave Shepard, Tetsuji Kuboyama, and Kilho Shin</i>	
Fast Community Discovery and Its Evolution Tracking in Time-Evolving Social Networks .....	13
<i>Yao Liu, Hong Gao, Xiaohui Kang, Qiao Liu, Ruijin Wang, and Zhiguang Qin</i>	
Correlation of Brand Mentions in Social Media and Web Searching Before and After Real Life Events: Phase Analysis of Social Media and Search Data for Super Bowl 2015 Commercials .....	21
<i>Partha Mukherjee and Bernard J. Jansen</i>	
Newsworthy Rumor Events: A Case Study of Twitter .....	27
<i>Armineh Nourbakhsh, Xiaomo Liu, Sameena Shah, Rui Fang, Mohammad Mahdi Ghassemi, and Quanzhi Li</i>	
Toponym Recognition in Social Media for Estimating the Location of Events .....	33
<i>Meryem Sagcan and Pinar Karagoz</i>	
Outlier Detection and Trend Detection: Two Sides of the Same Coin .....	40
<i>Erich Schubert, Michael Weiler, and Arthur Zimek</i>	
A Human-Machine Collaborative System for Identifying Rumors on Twitter .....	47
<i>Soroush Vosoughi and Deb Roy</i>	

## **Behavior Analysis, Modeling, and Steering**

Examining Botnet Behaviors for Propaganda Dissemination: A Case Study of ISIL’s Beheading Videos-Based Propaganda .....	51
<i>Samer Al-khateeb and Nitin Agarwal</i>	
Twitter Fingerprints as Active Authenticators .....	58
<i>Alex Brown and Myriam Abramson</i>	
Exploiting Class Bias for Discovery of Topical Experts in Social Media .....	64
<i>Iuliia Chepurina and Masoud Makrehchi</i>	
Population Behavior Analysis of Chinese University Students via Digital Campus Cards .....	72
<i>Shilin Fan, Ping Li, Ting Liu, and Yan Chen</i>	
Identifying Behavioral Characteristics in EGM Gambling Data Using Session Clustering .....	78
<i>Maria Gabriella Mosquera and Vlado Kešelj</i>	

## **Big Media Data: Understanding, Search, and Mining**

Generalized Learning of Neural Network Based Semantic Similarity Models and Its Application in Movie Search .....	86
<i>Xugang Ye, Zijie Qi, Xinying Song, Xiaodong He, and Dan Massey</i>	
Cross-Dataset Validation of Feature Sets in Musical Instrument Classification .....	94
<i>Patrick J. Donnelly and John W. Sheppard</i>	
Fast Keypoint Reduction for Image Retrieval by Accelerated Diverse Density Computation .....	102
<i>Toshikazu Wada and Yuichi Mukai</i>	
To Love or to Loathe: How is the World Reacting to China’s Rise? .....	108
<i>Yu Wang, Jianbo Yuan, and Jiebo Luo</i>	
Robust Local Outlier Detection .....	116
<i>Haizhou Du, Shengjie zhao, and Daqiang zhang</i>	

## **Biological Data Mining and Its Applications in Healthcare**

Understanding the Impact of Socio-Economic and Environmental Factors for Disease Outbreak in Developing Countries .....	124
<i>Zaheer Babar, Abdul Mannan, Faisal Kamiran, and Asim Karim</i>	
An Enumerative Biclustering Algorithm for DNA Microarray Data .....	132
<i>Haifa Ben Saber and Mourad Elloumi</i>	
Human Activity Recognition: From Controlled Lab Experiments to Competitive Live Evaluation .....	139
<i>Hristijan Gjoreski, Matjaž Gams, and Mitja Luštrek</i>	
Toward Comprehensive Attribution of Healthcare Cost Changes .....	146
<i>Dmitriy A. Katz-Rogozhnikov, Dennis Wei, Gigi Y. Yuen-Reed, Karthikeyan Natesan Ramamurthy, and Aleksandra Mojsilović</i>	
Predicting Clinical Phenotype Using OTU-Based Metagenome Representation .....	156
<i>Nathan LaPierre and Huzefa Rangwala</i>	

Activity Recognition Using Wrist-Worn Sensors for Human Performance Evaluation .....	164
<i>Minh Nguyen, Liyue Fan, and Cyrus Shahabi</i>	
Identifying Medical Terms Related to Specific Diseases .....	170
<i>Mihir Shekhar, Veera Raghavendra Chikka, Lini Thomas, Sunil Mandhan, and Kamalakar Karlapalem</i>	
Hierarchical Prescription Pattern Analysis with Symptom Labels .....	178
<i>Su-Jin Shin, Je-Yong Oh, Sungrae Park, Minki Kim, and Il-Chul Moon</i>	
Exploiting Social Media with Tensor Decomposition for Pharmacovigilance .....	188
<i>Christopher C. Yang and Haodong Yang</i>	
Mining a Weighted Heterogeneous Network Extracted from Healthcare-Specific Social Media for Identifying Interactions between Drugs .....	196
<i>Haodong Yang and Christopher C. Yang</i>	

## **Data Integration and Applications**

Methodology for Large-Scale Entity Resolution without Pairwise Matching .....	204
<i>Cheng Chen, Daniel Pullen, Reed H. Petty, and John R. Talburt</i>	
Context-Aware Approximate String Matching for Large-Scale Real-Time Entity Resolution .....	211
<i>Peter Christen and Ross W Gayler</i>	
Unsupervised Measuring of Entity Resolution Consistency .....	218
<i>Jeffrey Fisher and Qing Wang</i>	
Terra Populus: Integrated Data on Population and Environment .....	222
<i>Steven Ruggles, Tracy A. Kugler, Catherine A. Fitch, and David C. Van Riper</i>	
Building a National Perinatal Data Base without the Use of Unique Personal Identifiers .....	232
<i>Rainer Schnell and Christian Borgs</i>	
Multiresolution Mutual Information Method for Social Network Entity Resolution .....	240
<i>Cong Shi and Rong Duan</i>	
Named Entity Disambiguation Leveraging Multi-aspect Information .....	248
<i>Quanlong Zhang, Feng Li, Fang Wang, and Zhoujun Li</i>	

## **Data Mining for Educational Assessment and Feedback**

Temporal Models for Predicting Student Dropout in Massive Open Online Courses .....	256
<i>Mi Fei and Dit-Yan Yeung</i>	
Identifying Students' Mechanistic Explanations in Textual Responses to Science Questions with Association Rule Mining .....	264
<i>Yu Guo, Wanli Xing, and Hee-Sun Lee</i>	
Education, Learning and Information Theory .....	269
<i>Bryan Hooi, Hyun Ah Song, Evangelos Papalexakis, Rakesh Agrawal, and Christos Faloutsos</i>	
Including Content-Based Methods in Peer-Assessment of Open-Response Questions .....	273
<i>Oscar Luaces, Jorge Díez, Amparo Alonso, Alicia Troncoso, and Antonio Bahamonde</i>	

Using and Improving Coding Guides for and by Automatic Coding of PISA Short Text Responses .....	280
<i>Fabian Zehner, Frank Goldhammer, and Christine Sälzer</i>	

## **Data Mining for Services**

Adversarial Feature Selection .....	288
<i>Karan Kumar Budhraj and Tim Oates</i>	
Exploiting User Feedback for Expert Finding in Community Question Answering .....	295
<i>Xiang Cheng, Shuguang Zhu, Gang Chen, and Sen Su</i>	
J-Distance Discord: An Improved Time Series Discord Definition and Discovery Method .....	303
<i>Tian Huang, Yongxin Zhu, Yafei Wu, and Weiwei Shi</i>	
A Stochastic Game Theoretic Model for Expanding ATM Services .....	311
<i>Raja Rathnam Naidu Kanapaka and Raghu Kisore Neeliseti</i>	
Applying Data Mining Techniques to Direct Marketing: Challenges and Solutions .....	319
<i>Ying Li, Pavankumar Murali, Nan Shao, and Anshul Sheopuri</i>	
Why do Retailers End Price Promotions: A Study on Duration and Profit Effects of Promotion .....	328
<i>Zhen Li and Katsutoshi Yada</i>	
The Change of the Structure of Patient Satisfaction by Waiting Time, Length of Stay and Hospital Rebuilding in Japan .....	336
<i>Masumi Okuda, Akira Yasuda, and Shusaku Tsumoto</i>	
Identifying Employees for Re-skilling Using an Analytics-Based Approach .....	345
<i>Karthikeyan Natesan Ramamurthy, Moninder Singh, Michael Davis, J. Alex Kevern, Uri Klein, and Michael Peran</i>	
Adslot Mining for Online Display Ads .....	355
<i>Kazuki Taniguchi, Yuki Harada, and Nguyen Tuan Duc</i>	
Rule Mining Based on Nonmonotonic Rule Layers and Its Application to Medicine .....	365
<i>Shusaku Tsumoto and Shoji Hirano</i>	
Research Paper Recommendation Based on the Knowledge Gap .....	373
<i>Weidong Zhao, Ran Wu, Weihui Dai, and Yonghui Dai</i>	

## **Data Mining for Smart Cities**

Emergence of Regularities in the Stochastic Behavior of Human .....	381
<i>Marc Le Goc, Fabien Barthelot, and Eric Pascual</i>	
A Generic Energy Disaggregation Approach: What and When Electrical Appliances are Used .....	389
<i>Tengfei Liu, Xianghua Ding, and Ning Gu</i>	
Trajectory-Based Task Allocation for Reliable Mobile Crowd Sensing Systems .....	398
<i>Petar Mrazovic, Mihhail Matskin, and Nima Dokoohaki</i>	
Estimating Taxi Demand-Supply Level Using Taxi Trajectory Data Stream .....	407
<i>Dongxu Shao, Wei Wu, Shili Xiang, and Yu Lu</i>	
Valuating Queries for Data Trading in Modern Cities .....	414
<i>Ruiming Tang, Huayu Wu, Xiuqiang He, and Stéphane Bressan</i>	

Next Generation of Journey Planner in a Smart City .....	422
<i>Liang Yu, Dongxu Shao, and Huayu Wu</i>	
Learning Optimization Friendly Comfort Model for HVAC Model Predictive Control .....	430
<i>Yuxun Zhou, Dan Li, and Costas J. Spanos</i>	

## **Data Mining in Biomedical Informatics and Healthcare**

Affine and Regional Dynamic Time Warping .....	440
<i>Tsu-Wei Chen, Meena Abdelmaseeh, and Daniel Stashuk</i>	
Laplacian SVM Based Feature Selection Improves Medical Event Reports Classification .....	449
<i>Samah Fodeh, Andrea Benin, Perry Miller, Kyle Lee, Michele Koss, and Cynthia Brandt</i>	
Extensible Query Framework for Unstructured Medical Data—A Big Data Approach .....	455
<i>Sarmad Istephan and Mohammad-Reza Siadat</i>	
Applying Association Rule Mining to Semantic Data in the Lung Image Database Consortium .....	463
<i>Brendan Kennedy, Miguel Carrazza, Alex Rasin, Jacob Furst, and Daniela S. Raicu</i>	
A Content Analysis of Patient-Reported Medication Outcomes on Social Media .....	472
<i>Boshu Ru, Kimberly Harris, and Lixia Yao</i>	
Improving Mortality Prediction in Cardiovascular Risk Patients by Balancing Classes .....	480
<i>Emilio Soria-Olivas, José David Martín-Guerrero, Josep Redón, María Tellez-Plaza, and Joan Vila-Francés</i>	

## **Data Mining in Networks**

Link Prediction in Large Networks by Comparing the Global View of Nodes in the Network .....	485
<i>Mustafa Coşkun and Mehmet Koyutürk</i>	
StructMatrix: Large-Scale Visualization of Graphs by Means of Structure Detection and Dense Matrices .....	493
<i>Hugo Gualdrón, Robson L. F. Cordeiro, and Jose F. Rodrigues Jr.</i>	
Redundancy Reduction in Wireless Sensor Networks via Centrality Metrics .....	501
<i>Decebal Constantin Mocanu, Maria Torres Vega, and Antonio Liotta</i>	
Mining Unstable Communities from Network Ensembles .....	508
<i>Ahsanur Rahman, Steve Jan, Hyunju Kim, B. Aditya Prakash, and T. M. Murali</i>	
Detecting Anomalies in Time-Varying Networks Using Tensor Decomposition .....	516
<i>Anna Sapienza, André Panisson, Joseph Wu, Laetitia Gauvin, and Ciro Cattuto</i>	
Dispersion Based Similarity for Mining Similar Papers in Citation Network .....	524
<i>Saksham Singhal and Vikram Pudi</i>	
Efficient and Time Scale-Invariant Detection of Correlated Activity in Communication Networks .....	532
<i>Brian Thompson and James Abello</i>	
Email Engagement Segmentation Using Bipartite Graph Co-clustering .....	540
<i>Ketong Wang and Aaron Beach</i>	

Finding Hierarchical Communities in Complex Networks Using Influence-Guided Label Propagation .....	547
<i>Wenjun Wang and W. Nick Street</i>	

## **Data Science and Big Data Analytics**

A Study of Feature Selection of Magnetogram Complexity Features in an Imbalanced Solar Flare Prediction Data-Set .....	557
<i>Amani Al-Ghraibah, Laura E. Boucheron, and R. T. James McAteer</i>	
FPGA-Accelerated Hadoop Cluster for Deep Learning Computations .....	565
<i>Abdulrahman Alhamali, Nibal Salha, Raghid Morcel, Mazen Ezzeddine, Omar Hamdan, Haitham Akkary, and Hazem Hajj</i>	
Spatiotemporal Frequent Pattern Mining on Solar Data: Current Algorithms and Future Directions .....	575
<i>Berkay Aydin and Rafal Angryk</i>	
Unsupervised Learning Techniques for Detection of Regions of Interest in Solar Images .....	582
<i>Juan M. Banda and Rafal A. Angryk</i>	
Citation Prediction Using Diverse Features .....	589
<i>Harish S. Bhat, Li-Hsuan Huang, Sebastian Rodriguez, Rick Dale, and Evan Heit</i>	
Mining Massive Vector Data on Single Instruction Multiple Data Microarchitectures .....	597
<i>Christian Böhm and Claudia Plant</i>	
Stream Clustering: Efficient Kernel-Based Approximation Using Importance Sampling .....	607
<i>Radha Chitta, Rong Jin, and Anil K. Jain</i>	
Order Selection of Autoregressive Processes Using Bridge Criterion .....	615
<i>Jie Ding, Mohammad Noshad, and Vahid Tarokh</i>	
LeSiNN: Detecting Anomalies by Identifying Least Similar Nearest Neighbours .....	623
<i>Guansong Pang, Kai Ming Ting, and David Albrecht</i>	
A Deep Embedding Model for Co-occurrence Learning .....	631
<i>Yelong Shen, Ruoming Jin, Jianshu Chen, Xiaodong He, Jianfeng Gao, and Li Deng</i>	
Positive-Unlabeled Learning in the Face of Labeling Bias .....	639
<i>Noah Youngs, Dennis Shasha, and Richard Bonneau</i>	

## **DAVA: Data mining meets Visual Analytics at Big Data Era**

Visual Opinion Analysis of Threaded Discussions .....	646
<i>Yang Chen</i>	
Block-Organized Topology Visualization for Visual Exploration of Signed Networks .....	652
<i>Xianlin Hu, Leting Wu, Aidong Lu, and Xintao Wu</i>	
Fun in the Philippines: Automatic Identification and Sentiment Analysis of Tourism-Related Tweets .....	660
<i>Julia Camille L. Menchavez and Kurt Junshean P. Espinosa</i>	



Multi-sensor Visual Analytics Supported by Machine-Learning Models .....	668
<i>Geetika Sharma, Gautam Shroff, Aditeya Pandey, Brijendra Singh, Gunjan Sehgal, Kaushal Paneri, and Puneet Agarwal</i>	
OLAP Visual Analytics on Large Software Call Graphs with Hierarchical ChordMap .....	675
<i>Le Wang, Jiangtao Gong, and Lei Shi</i>	

## **Designing Safe and Secure Life on the Market of Data**

The Role of Context in Abductive Reasoning and Context Retrieval from Clauses—Towards IMDJ .....	680
<i>Akinori Abe</i>	
Proposal of LDA-Based Sentiment Visualization of Hotel Reviews .....	687
<i>Yu-Sheng Chen, Lieu-Hen Chen, and Yasufumi Takama</i>	
Extraction of Preference of Recipe Providers and Users on Recipe-Sharing Websites .....	694
<i>Mamoru Emoto</i>	
Estimating Contextual Relationships of Stakeholders in Scenarios Using DBpedia .....	698
<i>Teruaki Hayashi</i>	
Auctioning Data for Learning .....	706
<i>Iordanis Koutsopoulos, Aristides Gionis, and Maria Halkidi</i>	
Extended Goal Graph: A Support Tool for Discovering Conflicts among Stakeholders and Promoting Requirements Elicitation with Goal Orientation .....	714
<i>Noriyuki Kushiro and Takuro Shimizu</i>	
Shikake Data Market for Collaborative Shikake Creation .....	722
<i>Naohiro Matsumura and Hideaki Takeda</i>	
Knowledge-Based Circulation Growth Model: Applying a Data Marketplace to Concept Design .....	728
<i>Jun Nakamura and Masahiko Teramoto</i>	
Tangled String Diverted for Evaluating Stock Risks: A by Product of Innovators Marketplace on Data Jackets .....	734
<i>Yukio Ohsawa</i>	
Application of Applied KOTO-FRAME to the Five-Story Pagoda Aseismatic Mechanism .....	742
<i>Masahiko Teramoto and Jun Nakamura</i>	
Data Jacket Retrieval Based on Explicit Semantic Analysis .....	749
<i>Quexuan Zhang</i>	

## **Environmental Acoustic Representation and Data Mining**

Scattering Decomposition for Massive Signal Classification: From Theory to Fast Algorithm and Implementation with Validation on International Bioacoustic Benchmark .....	753
<i>Randall Balestrierio and Hervé Glotin</i>	
Compact Features for Birdcall Retrieval from Environmental Acoustic Recordings .....	762
<i>Xueyan Dong, Michael Towsey, Jinglan Zhang, and Paul Roe</i>	
Data Mining Sound Archives: A New Scalable Algorithm for Parallel-Distributing Processing .....	768
<i>Peter J. Dugan, Holger Klinck, John A. Zollweg, and Christopher W. Clark</i>	

JMesh—A Scalable Web-Based Platform for Visualization and Mining of Passive Acoustic Data .....	773
<i>Xavier Mouy, Pierre-Alain Mouy, David Hannay, and Tom Dakin</i>	
Sparse Coding for Efficient Bioacoustic Data Mining: Preliminary Application to Analysis of Whale Songs .....	780
<i>Joseph Razik, Hervé Glotin, Maia Hoeberechts, Yann Doh, and Sébastien Paris</i>	
The Navigation and Visualisation of Environmental Audio Using Zooming Spectrograms .....	788
<i>Michael W. Towsey, Anthony M. Truskinger, and Paul Roe</i>	
Computer-Assisted Sampling of Acoustic Data for More Efficient Determination of Bird Species Richness .....	798
<i>Liang Zhang, Michael Towsey, Jinglan Zhang, and Paul Roe</i>	

## **Interdisciplinary Workshop on Big Data**

Large Data Clustering Using Quadratic Programming: A Comprehensive Quantitative Analysis .....	806
<i>Alireza Chakeri and Lawrence O. Hall</i>	
Paradigmatic Clustering for NLP .....	814
<i>Julio Santisteban and Javier Tejada-Cárcamo</i>	
A Dirichlet Energy Criterion for Graph-Based Image Segmentation .....	821
<i>Dominique Zosso, Braxton Osting, and Stanley J. Osher</i>	

## **High Dimensional Data Mining**

Finding Subspace Clusters Using Ranked Neighborhoods .....	831
<i>Emin Aksehirli, Siegfried Nijssen, Matthijs van Leeuwen, and Bart Goethals</i>	
PaMPa-HD: A Parallel MapReduce-Based Frequent Pattern Miner for High-Dimensional Data .....	839
<i>Daniele Apiletti, Elena Baralis, Tania Cerquitelli, Paolo Garza, Pietro Michiardi, and Fabio Pulvirenti</i>	
HARAM: A Hierarchical ARAM Neural Network for Large-Scale Text Classification .....	847
<i>Fernando Benites and Elena Sapozhnikova</i>	
New Quality Indexes for Optimal Clustering Model Identification with High Dimensional Data .....	855
<i>Jean-Charles Lamirel and Pascal Cuxac</i>	
Pruned Simple Model Sets for Fast Exact Recovery of Image .....	863
<i>Basarab Matei and Younès Bennani</i>	
Fast LMNN Algorithm through Random Sampling .....	871
<i>Kaiyuan Wu and Zhiming Zheng</i>	

## **Information Analysis and Data Mining over Social Network**

Temporal Topic Inference for Trend Prediction .....	877
<i>Somayyeh Aghababaei and Masoud Makrehchi</i>	
Topic Detection Based on User Intention .....	885
<i>Lu Deng, Yong Quan, Jing Xu, Jiuming Huang, and Bin Zhou</i>	

Ranking-Based Multi-source Rating Aggregation with Social Contexts .....	892
<i>Lei Li, Qin Liang, and Guanfeng Liu</i>	
Dynamic Community Detection Algorithm Based on Incremental Identification .....	900
<i>Xiaoming Li, Bin Wu, Qian Guo, Xuelin Zeng, and Chuan Shi</i>	
Analyzing the Transferability of Collective Inference Models Across Networks .....	908
<i>Ransen Niu, Sebastian Moreno, and Jennifer Neville</i>	
Discovering Functional Communities in Social Media .....	917
<i>Brian Thompson, Linda Ness, David Shallcross, and Devasis Bassu</i>	
Joint Recovery and Representation Learning for Robust Correlation Estimation Based on Partially Observed Data .....	925
<i>Shupeng Wang, Xiao-Yu Zhang, Xiaochun Yun, and Guangjun Wu</i>	
Two Phase Based Spammer Detection in Weibo .....	932
<i>Xinhu Zheng, Jiamiao Wang, Fei Jie, and Lei Li</i>	
Core Network Based Multi-label Classification in Large-Scale Social Network Environments .....	940
<i>Zan Zhang, Hao Wang, Lei Li, and Guanfeng Liu</i>	
 <b>Intelligence and Security Informatics</b>	
Comparing SVD and SDAE for Analysis of Islamist Forum Postings .....	948
<i>N. Alsadhan and D. B. Skillicorn</i>	
Detecting Multipliers of Jihadism on Twitter .....	954
<i>Lisa Kaati, Enghin Omer, Nico Prucha, and Amendra Shrestha</i>	
Text Classification Models for Web Content Filtering and Online Safety .....	961
<i>Shuhua Liu and Thomas Fors</i>	
Identifying Key-Players in Online Activist Groups on the Facebook Social Network .....	969
<i>Mariam Nouh and Jason R. C. Nurse</i>	
Sentiment-Based Identification of Radical Authors (SIRA) .....	979
<i>Ryan Scrivens, Garth Davies, Richard Frank, and Joseph Mei</i>	
A Hybrid Learning from Multi-behavior for Malicious Domain Detection on Enterprise Network .....	987
<i>Liang Shi, Derek Lin, Chunsheng Victor Fang, and Yan Zhai</i>	
A Decision Support System for Crowd Control Using Agent-Based Modeling and Simulation .....	997
<i>Andrew J. Park, Herbert H. Tsang, Shawn Buckley, Hector C. Alzate Ramirez, and Valerie Spicer</i>	
Defending Suspected Users by Exploiting Specific Distance Metric in Collaborative Filtering Recommender Systems .....	1001
<i>Zhihai Yang and zhongmin Cai</i>	
Signed Directed Social Network Analysis Applied to Group Conflict .....	1007
<i>Q. Zheng, D. B. Skillicorn, and O. Walther</i>	
Constructing Topic Hierarchies from Social Media Data .....	1015
<i>Yuhao Zhang, Wenji Mao, and Daniel Zeng</i>	

## MobileDM & HuMoComp 2015

Mining Spatial-Temporal Semantic Trajectory Patterns from Raw Trajectories .....	1019
<i>Chien-Cheng Chen, Chia-Hsiang Kuo, and Wen-Chih Peng</i>	
How to Find Appropriate Automobile Exhibition Halls: Towards a Personalized Recommendation Service for Auto Show .....	1025
<i>Danhuai Guo, Yingqiu Zhu, Wei Xu, Yong Ge, Yuanchun Zhou, and Jianhui Li</i>	
A Method of Purchase Prediction Based on User Behavior Log .....	1031
<i>Dancheng Li, Guangming Zhao, Zhi Wang, Wenjia Ma, and Ying Liu</i>	
Impulse-Response Model for Human Behaviors Sequences .....	1040
<i>Houyi Li, Banghe Han, Ying Li, and Junming Li</i>	
Multi-Classes Feature Engineering with Sliding Window for Purchase Prediction in Mobile Commerce .....	1048
<i>Qiang Li, Maojie Gu, Keren Zhou, and Xiaoming Sun</i>	
Deep Convolutional Neural Network and Multi-view Stacking Ensemble in Ali Mobile Recommendation Algorithm Competition: The Solution to the Winning of Ali Mobile Recommendation Algorithm .....	1055
<i>Xiang Li, Suchi Qian, Furong Peng, Jian Yang, Xiaolin Hu, and Rui Xia</i>	
Lifting the Predictability of Human Mobility on Activity Trajectories .....	1063
<i>Xianming Li, Defu Lian, Xing Xie, and Guangzhong Sun</i>	
The Hierarchical Model to Ali Mobile Recommendation Competition .....	1070
<i>Suchi Qian, Furong Peng, Xiang Li, and Jianfeng Lu</i>	
Investigation of Simulated Annealing Cooling Schedule for Mobile Recommendations .....	1078
<i>Zeyang Ye, Keli Xiao, and Yuefan Deng</i>	
Purchase Behavior Prediction in M-Commerce with an Optimized Sampling Methods .....	1085
<i>Zhihao Yi, Danning Wang, Kai Hu, and Qiang Li</i>	
Profit Maximization Analysis Based on Data Mining and the Exponential Retention Model Assumption with Respect to Customer Churn Problems .....	1093
<i>Zhaojing Zhang, Regina Wang, Weihong Zheng, Shizhan Lan, Dong Liang, and Hao Jin</i>	

## Mobility Analytics from Spatial and Social Data

Observing Behaviors of Information Diffusion Models for Diverse Topics of Posts on VK .....	1098
<i>Aisylu Khairullina, JooYoung Lee, Gwan Jang, and Sung-Hyon Myaeng</i>	
Principal Component Analysis and Clustering Based Indoor Localization .....	1103
<i>Dong Liang, Jingkang Yang, Rui Xuan, Zhaojing Zhang, Zhifang Yang, and Kexin Shi</i>	
Trip Fare Estimation Study from Taxi Routing Behaviors and Localizing Traces .....	1109
<i>Ce Liu and Qiang Qu</i>	
Beyond Sights: Large Scale Study of Tourists' Behavior Using Foursquare Data .....	1117
<i>Ana Paula G. Ferreira, Thiago Henrique Silva, and Antonio Alfredo Ferreira Loureiro</i>	
Hospitality Exchange Services as a Source of Spatial and Social Data? .....	1125
<i>Rustam Tagiew, Dmitry I. Ignatov, and Radhakrishnan Delhibabu</i>	

Location Prediction via Social Contents and Behaviors: Location-Aware Behavioral LDA .....1131  
*Anna Tiginova, JooYoung Lee, and Sadegh Nobari*

Automatic City Region Analysis for Urban Routing .....1136  
*Kai Zhao, Mohan Prasath Chinnasamy, and Sasu Tarkoma*

## **Optimization Based Techniques for Emerging Data Mining Problems**

Similarity Learning for Product Recommendation and Scoring Using Multi-channel Data .....1143  
*Ifikhar Ahamath Burhanuddin, Payal Bajaj, Sumit Shekhar, Dipayan Mukherjee, Ashish Raj,  
and Aravind Sankar*

Optimized Swarm Search-Based Feature Selection for Text Mining in Sentiment Analysis .....1153  
*Simon Fong, Elisa Gao, and Raymond Wong*

Locality Sensitive Proximal Classifier with Consistency for Small Sample Size Problem .....1163  
*Yuan-Hai Shao, Zhen Wang, Chun-Na Li, and Nai-Yang Deng*

Alternating Direction Method of Multipliers for Nonparallel Support Vector Machines .....1171  
*Xin Shen, Lingfeng Niu, Yingjie Tian, and Yong Shi*

Large-Scale Linear Support Vector Ordinal Regression Solver .....1177  
*Yong Shi, Huadong Wang, and Lingfeng Niu*

Pedestrian Detection Using Privileged Information .....1185  
*Zhiquan Qi, Yingjie Tian, Lingfeng Niu, Fan Meng, Limeng Cui, and Yong Shi*

Matrix Factorization with Column L0-Norm Constraint for Robust Multi-subspace Analysis .....1189  
*Binghui Wang, Risheng Liu, Chuang Lin, and Xin Fan*

## **Practical Transfer Learning**

Adapting New Categories for Food Recognition with Deep Representation .....1196  
*Shuang Ao and Charles X. Ling*

CroRank: Cross Domain Personalized Transfer Ranking for Collaborative Filtering .....1204  
*Yunhui Guo, Xin Wang, and Congfu Xu*

Incremental Discriminant Learning for Heterogeneous Domain Adaptation .....1213  
*Peng Han and Xinxiao Wu*

Finding Event Videos via Image Search Engine .....1221  
*Han Wang and Xinxiao Wu*

## **Representation Learning for Semantic Data**

Learning Semantic Similarity for Very Short Texts .....1229  
*Cedric De Boom, Steven Van Canneyt, Steven Bohez, Thomas Demeester, and Bart Dhoedt*

Cross-Domain Recommendation via Tag Matrix Transfer .....1235  
*Zhou Fang, Sheng Gao, Bo Li, Juncen Li, and Jianxin Liao*

Music Mood Classification via Deep Belief Network .....1241  
*Juncen Li, Sheng Gao, Ning Han, Zhou Fang, and Jianxin Liao*

Matrix Plane Model: A Novel Measure of Word Co-occurrence and Application on Semantic Relatedness .....	1246
<i>Ji Qi and Yukio Ohsawa</i>	
A Unified Framework for Painting Classification .....	1254
<i>Babak Saleh and Ahmed Elgammal</i>	
Mining Individual Mobile User Behavior on Location and Interests .....	1262
<i>Junjie Yan, Yuanyuan Qiao, Jie Yang, and Sheng Gao</i>	

## **Sentiment Elicitation from Natural Text for Information Retrieval and Extraction**

Sentiment Polarity Classification Using Structural Features .....	1270
<i>Daniel Ansari</i>	
OntoSeg: A Novel Approach to Text Segmentation Using Ontological Similarity .....	1274
<i>Mostafa Bayomi, Killian Levacher, M. Rami Ghorab, and Séamus Lawless</i>	
Exploiting the Focus of the Document for Enhanced Entities' Sentiment Relevance Detection .....	1284
<i>Zvi Ben-Ami, Ronen Feldman, and Benjamin Rosenfeld</i>	
Improving Out-of-Domain Sentiment Polarity Classification Using Argumentation .....	1294
<i>Lucas Carstens and Francesca Toni</i>	
Lexical Resource for Medical Events: A Polarity Based Approach .....	1302
<i>Anupam Mondal, Iti Chaturvedi, Dipankar Das, Rajiv Bajpai, and Sivaji Bandyopadhyay</i>	
Production Estimation for Shale Wells with Sentiment-Based Features from Geology Reports .....	1310
<i>Bin Tong, Hiroaki Ozaki, Makoto Iwayama, Yoshiyuki Kobayashi, Sahu Anshuman, and Vennelakanti Ravigopal</i>	
An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis .....	1318
<i>Yun Wan and Qigang Gao</i>	
Towards Domain-Independent Opinion Target Extraction .....	1326
<i>Aleksander Wawer</i>	

## **Social Media Retrieval and Analysis**

A Multiple Classifier System for Classifying Life Events on Social Media .....	1332
<i>Paulo R. Cavalin, Luis G. Moyano, and Pedro P. Miranda</i>	
A Two-Stage Learning Method for Response Prediction .....	1336
<i>Kuan-Hsi Chen, Zih-Yun Ting, Jia-Ying Shen, Yuh-Jyh Hu, and Tyne Liang</i>	
Inferring User Interests on Social Media from Text and Images .....	1342
<i>Yagmur Gizem Cinar, Susana Zoghbi, and Marie-Francine Moens</i>	
Buzzer Detection and Sentiment Analysis for Predicting Presidential Election Results in a Twitter Nation .....	1348
<i>Mochamad Ibrahim, Omar Abdillah, Alfan F. Wicaksono, and Mirna Adriani</i>	
OLLDA: A Supervised and Dynamic Topic Mining Framework in Twitter .....	1354
<i>Shatha Jaradat, Nima Dokoohaki, and Mihhail Matskin</i>	

Towards a Context-Aware Music Recommendation Approach: What is Hidden in the Playlist Name? .....	1360
<i>Martin Pichl, Eva Zangerle, and Günther Specht</i>	
Towards Personalizing Classical Music Recommendations .....	1366
<i>Markus Schedl</i>	
A Novel Approach for Generating Personalized Mention List on Micro-Blogging System .....	1368
<i>Ge Zhou, Lu Yu, Chu-Xu Zhang, Chuang Liu, Zi-Ke Zhang, and Jianlin Zhang</i>	

## **Social Multimedia Data Mining**

Grandmaster: Interactive Text-Based Analytics of Social Media .....	1375
<i>Nathan Fabian, Warren Davis, Elaine M. Raybourn, Kiran Lakkaraju, and Jon Whetzel</i>	
Home Location Inference from Sparse and Noisy Data: Models and Applications .....	1382
<i>Tianran Hu, Jiebo Luo, Henry Kautz, and Adam Sadilek</i>	
Distributed Representations for Content-Based and Personalized Tag Recommendation .....	1388
<i>Saurabh Kataria and Arvind Agarwal</i>	
Optimizing the Popularity of Twitter Messages through User Categories .....	1396
<i>Rupert Lemahieu, Steven Van Canneyt, Cedric De Boom, and Bart Dhoedt</i>	
Towards Automatic Pharmacovigilance: Analysing Patient Reviews and Sentiment on Oncological Drugs .....	1402
<i>Arpit Mishra, Ankit Malviya, and Sanchit Aggarwal</i>	
Beyond Classification: Latent User Interests Profiling from Visual Contents Analysis .....	1410
<i>Longqi Yang, Cheng-Kang Hsieh, and Deborah Estrin</i>	

## **Spatial and Spatio-Temporal Data Mining**

Classification of Temporal and Sequential Data Using Bag-of-Subsequences Features .....	1417
<i>Shin Ando and Yutaro Suka</i>	
Efficient Distance-Based Gestural Pattern Mining in Spatiotemporal 3D Motion Capture Databases .....	1425
<i>Christian Beecks, Marwan Hassani, Florian Obeloer, and Thomas Seidl</i>	
Unsupervised Spatial Data Mining for Smart Homes .....	1433
<i>Kevin Bouchard, Dany Fortin-Simard, Jeremy Lapalu, Sebastien Gaboury, Abdenour Bouzouane, and Bruno Bouchard</i>	
Learning the Number of Autoregressive Mixtures in Time Series Using the Gap Statistics .....	1441
<i>Jie Ding, Mohammad Noshad, and Vahid Tarokh</i>	
Geographically Robust Hotspot Detection: A Summary of Results .....	1447
<i>Emre Eftelioglu, Xun Tang, and Shashi Shekhar</i>	
Extraction of Highly Correlated Temporal Event Cluster Recurrence from Spatiotemporal Data .....	1457
<i>Rie Honda and Keita Mori</i>	
Flood Prediction and Mining Influential Spatial Features on Future Flood with Causal Discovery .....	1462
<i>Piraporn Jangyodsuk, Dong-Jun (DJ) Seo, Ramez Elmasri, and Jean Gao</i>	

Specialist Experts for Prediction with Side Information .....	1470
<i>Yuri Kalnishkan, Dmitry Adamskiy, Alexey Chernov, and Tim Scarfe</i>	
Characterization and Classification of Vegetation Canopy Structure and Distribution within the Great Smoky Mountains National Park Using LiDAR .....	1478
<i>Jitendra Kumar, Jon Weiner, William W. Hargrove, Steven P. Norman, Forrest M. Hoffman, and Doug Newcomb</i>	
<b>PhD Forum</b>	
Accurate Classification of Biological Data Using Ensembles .....	1486
<i>Manju Bhardwaj, Debasis Dash, and Vasudha Bhatnagar</i>	
Latent Graph Inference and Validation .....	1494
<i>Ivan Brugere</i>	
Prediction of Long-Lead Heavy Precipitation Events Aided by Machine Learning .....	1496
<i>Yahui Di</i>	
Selecting Machine Learning Algorithms Using Regression Models .....	1498
<i>Tri Doan and Jugal Kalita</i>	
Modeling of Geographic Dependencies for Real Estate Ranking on Site Selection .....	1506
<i>Yanjie Fu and Hui Xiong</i>	
Subspace Model Based Discriminative Instances Selection for Weakly Supervised Object Detection .....	1514
<i>Qiaoying Huang, Xiaofeng Zhang, Kui Jia, Xishuang Han, and Yunming Ye</i>	
Learning Spatial Decision Trees for Land Cover Mapping .....	1522
<i>Zhe Jiang</i>	
Building Predictive Models for Noisy and Heterogeneous Data: An Application in Global Monitoring of Inland Water Dynamics .....	1530
<i>Anuj Karpatne and Vipin Kumar</i>	
Enhancing Network Security by Software Vulnerability Detection Using Social Media Analysis Extended Abstract .....	1532
<i>Dennis Kergl</i>	
Clustering Evolving Batch System Jobs for Online Anomaly Detection .....	1534
<i>Eileen Kuehn</i>	
Online Change Detection Algorithm for Noisy Time-Series: An Application Tonear-Real Time Burned Area Mapping .....	1536
<i>Xi C. Chen, Vipin Kumar, and James H Faghmous</i>	
Secure Learning and Mining in Adversarial Environments [Extended Abstract] .....	1538
<i>Bo Li</i>	
Crowdsourcing Based API Search via Leveraging Twitter Lists Information .....	1540
<i>Tingting Liang, Liang Chen, Haochao Ying, Zibin Zheng, and Jian Wu</i>	
Understanding Deep Networks with Gradients .....	1548
<i>Henry Z. Lo and Wei Ding</i>	



Probabilistic Models for Fine-Grained Truth Discovery from Crowdsourced Data .....	1556
<i>Fenglong Ma and Jing Gao</i>	
Revenue-Optimized Webpage Recommendation .....	1558
<i>Chong Wang, Achir Kalra, Cristian Borcea, and Yi Chen</i>	
Contrast Pattern Based Methods for Visualizing and Predicting Spatiotemporal Events .....	1560
<i>Dawei Wang</i>	
Enhancing Stock Price Prediction with a Hybrid Approach Based Extreme Learning Machine .....	1568
<i>Feng Wang, Yongquan Zhang, Hang Xiao, Li Kuang, and Yi Lai</i>	
Deep Learning for Image Retrieval: What Works and What Doesn't .....	1576
<i>Huafeng Wang, Yehe Cai, Yanxiang Zhang, Haixia Pan, Weifeng Lv, and Hao Han</i>	
Sentiment Analysis for Social Media Images .....	1584
<i>Yilin Wang and Baoxin Li</i>	
Reporting L Most Favorite Objects in Uncertain Databases with Probabilistic Reverse Top-k Queries .....	1592
<i>Guoqing Xiao, Kenli Li, and Keqin Li</i>	
Multi-source Information Trustworthiness Analysis .....	1600
<i>Houping Xiao and Jing Gao</i>	
Context Suggestion: Solutions and Challenges .....	1602
<i>Yong Zheng</i>	

## **Demos**

IntelligShop: Enabling Intelligent Shopping in Malls through Location-Based Augmented Reality .....	1604
<i>Aditi Adhikari, Vincent W. Zheng, Hong Cao, Miao Lin, Yuan Fang, and Kevin Chen-Chuan Chang</i>	
StreamDM: Advanced Data Mining in Spark Streaming .....	1608
<i>Albert Bifet, Silviu Maniu, Jianfeng Qian, Guangjian Tian, Cheng He, and Wei Fan</i>	
A Discovery System for Finding High-Value Homes .....	1612
<i>Yanjie Fu and Hui Xiong</i>	
Large-Scale Unusual Time Series Detection .....	1616
<i>Rob J. Hyndman, Earo Wang, and Nikolay Laptev</i>	
TeGViz: Distributed Tera-Scale Graph Generation and Visualization .....	1620
<i>ByungSoo Jeon, Inah Jeon, and U. Kang</i>	
Near Real-Time Service Monitoring Using High-Dimensional Time Series .....	1624
<i>Shwetabh Khanduja, Vinod Nair, S. Sundararajan, Ameya Raul, Ajesh Babu Shaj, and Sathiya Keerthi</i>	
Trend-MC: A Melody Composer by Constructing from Frequent Trend-Based Patterns .....	1628
<i>Cheng Long, Raymond Chi-Wing Wong, and Raymond Ka Wai Sze</i>	
Social Answer: A System for Finding Appropriate Sites for Questions in Social Media .....	1632
<i>Harsh Dani, Fred Morstatter, Xia Hu, Zhen Yang, and Huan Liu</i>	
TextServer: Cloud-Based Multilingual Natural Language Processing .....	1636
<i>Lluís Padró and Jordi Turmo</i>	

MERLIN—A Tool for Multi-party Privacy-Preserving Record Linkage .....	1640
<i>Thilina Ranbaduge, Dinusha Vatsalan, and Peter Christen</i>	
Automatically Generate a Flat Mining Table with Dataconda .....	1644
<i>Michele Samorani</i>	
Real-Time Crisis Mapping Using Language Distribution .....	1648
<i>Justin Sampson, Fred Morstatter, Reza Zafarani, and Huan Liu</i>	
Influence Visualization of Scientific Paper through Flow-Based Citation Network Summarization .....	1652
<i>Yue Su, Sibai Sun, Yuan Xuan, and Lei Shi</i>	
UP-Miner: A Utility Pattern Mining Toolbox .....	1656
<i>Vincent S. Tseng, Cheng-Wei Wu, Jun-Han Lin, and Philippe Fournier-Viger</i>	
Cross-Site Virtual Social Network Construction .....	1660
<i>Chenhao Xie, Deqing Yang, Jingrui He, and Yanghua Xiao</i>	
Discovering Anomalies and Root Causes in Applications via Relevant Fields Analysis .....	1664
<i>Yuchen Zhao, Arjun Iyer, and Ariel Smoliar</i>	
CARSKit: A Java-Based Context-Aware Recommendation Engine .....	1668
<i>Yong Zheng, Bamshad Mobasher, and Robin Burke</i>	
Mining Predictive Redescriptions with Trees .....	1672
<i>Tetiana Zinchenko, Esther Galbrun, and Pauli Miettinen</i>	
<b>ICDM Contest</b>	
Machine Learning Approach to Identify Users Across Their Digital Devices .....	1676
<i>Thakur Raj Anand and Oleksii Renov</i>	
Recovering Cross-Device Connections via Mining IP Footprints with Ensemble Learning .....	1681
<i>Xuezhi Cao, Weiyue Huang, and Yong Yu</i>	
Cross-Device Consumer Identification .....	1687
<i>Girma Kejela and Chunming Rong</i>	
Connecting Devices to Cookies via Filtering, Feature Engineering, and Boosting .....	1690
<i>Michael Sungjun Kim, Jiwei Liu, Xiaozhou Wang, and Wei Yang</i>	
Multi-layer Classification: ICDM 2015 Drawbridge Cross-Device Connections Competition .....	1695
<i>Mark Landry, Sudalai Rajkumar S, and Robert Chong</i>	
Cross-Device Tracking: Matching Devices and Cookies .....	1699
<i>Roberto Díaz-Morales</i>	
AFFM: Auto Feature Engineering in Field-Aware Factorization Machines for Predictive Analytics .....	1705
<i>Lars Ropeid Selsaas, Bikash Agrawal, Chunming Rong, and Thomasz Wiktorski</i>	
Learning to Rank for Cross-Device Identification .....	1710
<i>Jeremy Walthers</i>	

## Author Index