Table of Contents

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

CLOUDMINE + DMIoT

CLOUDMINE — Data Mining Systems and their Applications on the Cloud

Database Integrated Analytics Using R: Initial Experiences with SQL-Server + R .................................................................1
Josep Ll. Berral and Nicolas Poggi

Parallelized Frequent Item Set Mining Using a Tall and Skinny Matrix ..................................................................................8
D. Pooja Janakiram

A Probabilistic View of Neighborhood-Based Recommendation Methods .............................................................................14
Jun Wang and Qiang Tang

Discovering Multi-type Correlated Events with Time Series for Exception Detection of Complex Systems .................................................21
Peng Xun, Pei-Dong Zhu, Cun-Lu Li, and Hao-Yang Zhu

DMIoT — Data Mining for Internet of Things

Detecting Performance Degradation of Software-Intensive Systems in the Presence of Trends and Long-Range Dependence ..............................................................29
Alexey Artemov and Evgeny Burnaev

Scalable Online-Offline Stream Clustering in Apache Spark ...........................................................................................................37
Omar Backhoff and Eirini Ntoutsi

Distributed Mining and Modeling of Dynamic Lead-Lag Relations in Evolving Entities ........................................................................45
Tian Guo, Jean-Paul Calbimonte, and Karl Aberer
Event Detection for Urban Dynamic Data Streams ..........................................................53

Septimiu Nechifor, Ioana Stefan, Marten Fischer, and Dan Puiu

**DaMNet — Data Mining in Networks**

Segmenting Sequences of Node-Labeled Graphs ..........................................................61

Sorour E. Amiri, Liangzhe Chen, and B. Aditya Prakash

Inference of Partial Canonical Correlation Networks with Application to Stock Market Portfolio Selection ..........................................................69

Gregory Breard and Natallia Katenka

Overlapping Community Detection by Local Decentralised Vertex-Centred Process ..................................................77

Maël Canu, Marie-Jeanne Lesot, and Adrien Revault D’Allonnes

Query-Based Evolutionary Graph Cuboid Outlier Detection ......................................85

Ayushi Dalmia, Manish Gupta, and Vasudeva Varma

Vertex Centric Asynchronous Belief Propagation Algorithm for Large-Scale Graphs ...........................................93

Gabriel P. Gimenes, Hugo Gualdron, and Jose F. Rodrigues-Jr

Text Network Exploration via Heterogeneous Web of Topics ....................................99

Junxian He, Ying Huang, Changfeng Liu, Jiaming Shen, Yuting Jia, and Xinbing Wang

Classification of Normal and Pathological Brain Networks Based on Similarity in Graph Partitions ..........................................................107

Anvar Kurmukov, Yulia Dodonova, and Leonid Zhukov

Finding Heaviest k-Subgraphs and Events in Social Media .......................................113

Matthaios Letsios, Oana Denisa Balalau, Maximilian Danisch, Emmanuel Orsini, and Mauro Sozio

Knowledge Graph Constraints for Multi-label Graph Classification ........................121

Martin Ringsquandl, Steffen Lamparter, Ingo Thon, Raffaello Lepratti, and Peer Kröger

Predicting User Roles in Social Networks Using Transfer Learning with Feature Transformation ..........................................................128

Jun Sun, Jérôme Kunegis, and Steffen Staab

**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

Javier Fernández, Daniel Medina, Antonio Gómez, Marta Arias, and Ricard Gavaldá

Competition-Wide Evaluation of Individual and Team Movements in Soccer ...............144

László Gyarmati and Mohamed Hefeeda

Football Market Strategies: Think Locally, Trade Globally ........................................152

Giulio Rossetti and Vincenzo Caproni
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characterization of In-season Elite Football Trainings by GPS Features: The Identity</td>
<td>160</td>
</tr>
<tr>
<td>Alessio Rossi, Enrico Perri, Athos Trecroci, Marco Savino, Giampietro Alberti, and Marcello Fedon Iaia</td>
<td></td>
</tr>
<tr>
<td>Towards Data-Driven Football Player Assessment</td>
<td>167</td>
</tr>
<tr>
<td>Rade Stanojevic and László Gyarmati</td>
<td></td>
</tr>
<tr>
<td>Stick Must Fall: Using Machine Learning to Predict Human Error in Virtual Balancing Task</td>
<td>173</td>
</tr>
<tr>
<td>Irina Zgonnikova, Arkady Zgonnikov, and Shigeru Kanemoto</td>
<td></td>
</tr>
<tr>
<td><strong>DINA — Data Integration and Applications</strong></td>
<td>178</td>
</tr>
<tr>
<td>Towards Information Profiling: Data Lake Content Metadata Management</td>
<td></td>
</tr>
<tr>
<td>Ayman Alserafi, Alberto Abelló, Oscar Romero, and Toon Calders</td>
<td></td>
</tr>
<tr>
<td>Historical Data Integration: A Study of WWI Canadian Soldiers</td>
<td>186</td>
</tr>
<tr>
<td>Luiza Antonie, Harshavardhan Gadgil, Gary Grewal, and Kris Inwood</td>
<td></td>
</tr>
<tr>
<td>Holistic Entity Clustering for Linked Data</td>
<td>194</td>
</tr>
<tr>
<td>Markus Nentwig, Anika Groß, and Erhard Rahm</td>
<td></td>
</tr>
<tr>
<td>Cluster-Based Sorted Neighborhood for Efficient Duplicate Detection</td>
<td>202</td>
</tr>
<tr>
<td>Ahmad Samiei and Felix Naumann</td>
<td></td>
</tr>
<tr>
<td>Integrated Analysis of Research Publications and Patents for Strategic Decision Making</td>
<td>210</td>
</tr>
<tr>
<td>Nidhi Saraswat, Lipika Dey, Ishan Verma, and Hemant Gupta</td>
<td></td>
</tr>
<tr>
<td>Randomized Response and Balanced Bloom Filters for Privacy Preserving Record Linkage</td>
<td>218</td>
</tr>
<tr>
<td>Rainer Schnell and Christian Borgs</td>
<td></td>
</tr>
<tr>
<td>A Probabilistic Address Parser Using Conditional Random Fields and Stochastic Regular Grammar</td>
<td>225</td>
</tr>
<tr>
<td>Minlue Wang, Valeriia Haberland, Amos Yeo, Andrew Martin, John Howroyd, and J. Mark Bishop</td>
<td></td>
</tr>
<tr>
<td><strong>DMBIH — Data Mining in Biomedical Informatics and Healthcare</strong></td>
<td>233</td>
</tr>
<tr>
<td>Construction of a Semi-Naive Model to Predict Early Readmission of COPD Patients by Using Quality Care Information</td>
<td></td>
</tr>
<tr>
<td>Pablo Bermejo, José A. Gámez, José M. Puerta, Marco A. Esquivias, and Pedro J. Tárraga</td>
<td>241</td>
</tr>
<tr>
<td>Medical Image Denoising Using Convolutional Denoising Autoencoders</td>
<td></td>
</tr>
<tr>
<td>Lovedeep Gondara</td>
<td></td>
</tr>
</tbody>
</table>
Machine Learning for Modeling the Biomechanical Behavior of Human Soft Tissue .........................................................247
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

Applying Conditional Independence Maps to Improve Sepsis Prognosis .................................................................254
Carles Morales, Vicent Ribas, and Alfredo Vellido

Ligand-Based Virtual Screening with Co-regularised Support Vector Regression .........................................................261
Katrin Ullrich, Michael Kamp, Thomas Gärtner, Martin Vogt, and Stefan Wrobel

DMCS — Data Mining for Cyber Security

Uncertainty Aware Clustering for Behaviour in Enterprise Networks ..................................................................................269
Maha Bakoben, Niall Adams, and Anthony Bellotti

One-Class SVM with Privileged Information and Its Application to Malware Detection ...............................................................273
Evgeny Burnaev and Dmitry Smolyakov

Adaptive Threshold Selection for Trust-Based Detection Systems ................................................................................281
Younghun Chae, Natallia Katenka, and Lisa Dipippo

The Time-Varying Dependency Patterns of NetFlow Statistics ............................................................................................288
Alexander Gibberd, Marina Evangelou, and James D. B. Nelson

Learning a Secure Classifier against Evasion Attack ........................................................................................................295
Zeinab Khorshidpour, Sattar Hashemi, and Ali Hamzeh

Detection of Cyber-Physical Faults and Intrusions from Physical Correlations ...............................................................303
Andrey Y. Lokhov, Nathan Lemons, Thomas C. McAndrew, Aric Hagberg, and Scott Backhaus

Correlation-Based Streaming Anomaly Detection in Cyber-Security ................................................................................311
Jordan Noble and Niall M. Adams

Smart Phone User Behaviour Characterization Based on Autoencoders and Self Organizing Maps ........................................................319
Deepthi Rajashekar, A. Nur Zincir-Heywood, and Malcolm I. Heywood

EdgeCentric: Anomaly Detection in Edge-Attributed Networks ......................................................................................327
Neil Shah, Alex Beutel, Bryan Hooi, Leman Akoglu, Stephan Gummemann, Disha Makhija, Mohit Kumar, and Christos Faloutsos

Authenticating Web User’s Identity through Browsing Sequences Modeling ..................................................................335
Peihai Zhao, Chungang Yan, and Changjun Jiang
DMED I — Data Mining in Emerging Domains I

DWA — Data Wrangling Automation

Using Machine Learning to Accelerate Data Wrangling .................................................................343
  Shilpi Ahuja, Mary Roth, Rashmi Gangadharaiah, Peter Schwarz, and Rafael Bastidas

Mining the Dark Web: Drugs and Fake Ids ........................................................................350
  Andres Baravalle, Mauro Sanchez Lopez, and Sin Wee Lee

Toward Representation Independent Analytics over Structured Data ........................................357
  Yodsawalai Chodpathumwan, Jose Picado, Arash Termehchy, Alan Fern, and Yizhou Sun

MDDM — Maritime Domain Data Mining

Mining Vessel Tracking Data for Maritime Domain Applications ..............................................361
  Alfredo Alessandrini, Marlene Alvarez, Harm Greidanus, Vincenzo Gammieri,
  Virginia Fernandez Arguedas, Fabio Mazzarella, Carlos Santamaria, Mattia Stasolla,
  Dario Tarchi, and Michele Vespe

On the Interest of Data Mining for an Integrity Assessment of AIS Messages ..........................368
  Clément Iphar, Aldo Napoli, and Cyril Ray

Scalable and Distributed Sea Port Operational Areas Estimation from AIS Data ......................374
  Leonardo M. Millefiori, Dimitrios Zissis, Luca Cazzanti, and Gianfranco Arcieri

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
  M. Abdallatif, S. Schramm, and J. Götze

Online Outlier Detection of Energy Data Streams Using Incremental and Kernel PCA
  Algorithms .........................................................................................................................................390
  Jeremiah D. Deng

Adaptive Time Series Forecasting of Energy Consumption Using Optimized Cluster Analysis .........................................................................................................................398
  Peter Laurinec, Marek Lóderer, Petra Vrablecová, Mária Lucká, Viera Rozinajová,
  and Anna Bou Ezzeddine

Structural Combination of Neural Network Models ..................................................................406
  Juan Rendon and Lilian M. de Menezes

Incremental Time Series Prediction Using Error-Driven Informed Adaptation .......................414
  Petra Vrablecová, Viera Rozinajová, and Anna Bou Ezzeddine
DMiP — Data Mining in Politics

Analysing Political Opinions Using Redescription Mining ................................................................. 422

Esther Galbrun and Pauli Miettinen

Leave or Remain? Deciphering Brexit Deliberations on Twitter .......................................................... 428

Apalak Khatua and Aparup Khatua

Change-Point Analysis of the Public Mood in UK Twitter during the Brexit Referendum .................. 434

Thomas Lansdall-Welfare, Fabon Dzogang, and Nello Cristianini

Structural Patterns in the Rise of Germany’s New Right on Facebook .............................................. 440

Sebastian Schelter, Felix Biessmann, Malisa Zobel, and Nedelina Teneva

DMHAA — Data Mining in Human Activity Analysis

Weakly Supervised Hand Pose Recovery with Domain Adaptation by Low-Rank Alignment ................ 446

Chaoqun Hong, Jun Yu, Rongsheng Xie, and Dapeng Tao

Link Prediction in the Twitter Mention Network: Impacts of Local Structure and Similarity of Interest ................................................................. 454

Hadrien Hours, Eric Fleury, and Márton Karsai

Gender Classification by Deep Learning on Millions of Weakly Labelled Images ............................. 462

Sen Jia, Thomas Lansdall-Welfare, and Nello Cristianini

Real-Time Top-View People Counting Based on a Kinect and NVIDIA Jetson TK1 Integrated Platform ................................................................. 468

Guangqin Li, Peng Ren, Xinrong Lyu, and He Zhang

Online Semi-Supervised Multi-task Distance Metric Learning .......................................................... 474

Ya Li and Dacheng Tao

Local Structure Preserving Using Manifold Regularization and Pairwise Constraints for Action Recognition ................................................................. 480

Xueqi Ma, Zhengyang Cao, and Weifeng Liu

Stereotypical Motor Movement Detection in Dynamic Feature Space .............................................. 487

Nastaran Mohammadian Rad, Seyed Mostafa Kia, Calogero Zarbo, Giuseppe Jurman, Paola Venuti, and Cesare Furlanello

Robust Automated Human Activity Recognition and Its Application to Sleep Research .................. 495


Sequence-to-Sequence Model with Attention for Time Series Classification ....................................... 503

Yujin Tang, Jianfeng Xu, Kazunori Matsumoto, and Chihiro Ono
DMS — Data Mining for Service

Risk-Aware Dynamic Reserve Prices of Programmatic Guarantee in Display
Bowei Chen .................................................................511

A Novel Method for Assessing Event Impacts on Event-Driven Time Series
Lianhua Chi, Saket Sathe, Bo Han, and Yun Wang ...........................519

Temporal Data Mining on the Stay Time of Outpatients and Treatment Processes
Shoji Hirano .................................................................527

A Deep Learning Approach for the Prediction of Retail Store Sales
Yuta Kaneko and Katsutoshi Yada ..................................................531

Website Classification Using Latent Dirichlet Allocation and Its Application for Internet Advertising
Sotaro Katsumata, Eiji Motohashi, Akihiro Nishimoto, and Eiji Toyosawa .................................538

Personal Identification with Face and Voice Features Extracted through Kinect Sensor
Eisuke Kita, Yi Zuo, Fumiya Saito, and Xuanang Feng .................................545

Data Mining Based Product Marketing Technique for Banking Products
Merve Mitik, Ozan Korkmaz, Pinar Karagoz, Ismail Hakki Toroslu, and Ferhat Yucel ..................552

Interactive Independent Topic Analysis for Service
Takahiro Nishigaki, Katsumi Nitta, and Takashi Onoda .................................560

Activity Detection from Email Meta-Data Clustering
Mayur Patidar, Shaurya Rohatgi, Ashish Chaudhary, Mahesh P. Singh,
Puneet Agarwal, and Gautam Shroff ..................................................568

Feature Selection in Environments with Limited Voluntary Information Sharing
Nicholay Topin, Karan K. Budhraja, and Tim Oates ........................................576

Domain Driven Forecasting: Applying Theory Into Practice
Praveen Venkateswaran, Vikrant Shimpi, Maitreya Natu, and Vaishali Sadaphal .................................584

Multiscale Analysis of Supply Network at Central Region in Japan
Yi Zuo and Yuya Kajikawa .................................................................592

DSBDA — Data Science and Big Data Analytics

Visual Context Learning with Big Data Analytics
Mayanka Chandrashekar and Yuyung Lee ........................................600

Finding Large Matchings in Semi-Streaming
Hossein Esfandiari, MohammadTaghi Hajiaghayi, and Morteza Monemizadeh .................................608

Effective and Unsupervised Fractal-Based Feature Selection for Very Large Datasets:
Removing Linear and Non-linear Attribute Correlations
Antonio C. Fraideinberze, José F. Rodrigues Jr., and Robson L. F. Cordeiro .................................615
Tovel: Distributed Graph Clustering for Word Sense Disambiguation ................................................................. 623
  Alessio Guerrieri, Fatemeh Rahimian, Sarunas Girdzijauskas, and Alberto Montresor

Will I Get in? Modeling the Graduate Admission Process for American Universities ............................................. 631
  Narender Gupta, Aman Sawhney, and Dan Roth

Selecting the Top-k Discriminative Features Using Principal Component Analysis ........................................... 639
  Aminata Kane and Nematollaah Shiri

Score Look-Alike Audiences .................................................................................................................................. 647
  Qiang Ma, Eeshan Wagh, Jiayi Wen, Zhen Xia, Robert Ormandi, and Datong Chen

Designing Sketches for Similarity Filtering ........................................................................................................... 655
  Vladimir Mic, David Novak, and Pavel Zezula

Tracking Language Mobility in the Twitter Landscape .............................................................................................. 663
  Izabela Moise, Edward Gaere, Ruben Merz, Stefan Koch, and Evangelos Pournaras

Relative Label Encoding for the Prediction of Airline Passenger Nationality .......................................................... 671
  Alejandro Mottini and Rodrigo Acuna-Agost

WEFEST: Word Embedding Feature Extension for Short Text Classification ....................................................... 677
  Lei Sang, Fei Xie, Xiaoqian Liu, and Xindong Wu

Knowledge Representation Learning via Dynamic Relation Spaces .............................................................................. 684
  Zhen Tan, Xiang Zhao, Yang Fang, Weidong Xiao, and Jiuyang Tang

Steady Patterns ........................................................................................................................................................... 692
  Willy Ugarte, Alexandre Termier, and Miguel Santana

Mining Effective Subsequences with Application in Marketing Attribution .......................................................... 700
  Zi Yin, Ying Li, Pietro Mazzoleni, and Yuanyuan Shen

**HDM — High Dimensional Data Mining**

Random Projection Clustering on Streaming Data ...................................................................................................... 708
  Lee A. Carraher, Philip A. Wilsey, Anindya Moitra, and Sayantan Dey

An Empirical Analysis of Hubness in Unsupervised Distance-Based Outlier Detection .................................... 716
  Arthur Flexer

Topic Extraction Method from Millions of Tweets Based on Fast Feature Selection Technique CWC .............................. 724
  Takako Hashimoto, Dave Shepard, Tetsuji Kuboyama, and Kilho Shin

Regression on High-Dimensional Inputs .................................................................................................................. 732
  Alexander Kuleshov and Alexander Bernstein

Clustering Based on MultiView Diffusion Maps ......................................................................................................... 740
  Ofir Lindenbaum, Arie Yeredor, and Amir Averbuch

Aggregating Tree for Searching in Billion Scale High Dimensional Data .............................................................. 748
  Shicong Liu, Junru Shao, and Hongtao Lu
MoDAT — Data Market for Co-Evolution of Sciences and Business
Method for Extraction of Purchase Behavior and Product Character Using Dynamic
Topic Model ............................................................. 778
Mamoru Emoto
Inferring Variable Labels Considering Co-occurrence of Variable Labels in Data
Jackets ........................................................................ 783
Teruaki Hayashi
The Prominent Role of Probe Data and Big Data in Modern Technology ........................................ 788
Hideki Isetani
ID-Link, an Enabler for Medical Data Marketplace .............................................................. 792
Ryuji Ito
Predicting the Outcome of Startups: Less Failure, More Success .............................................. 798
Amar Krishna, Ankit Agrawal, and Alok Choudhary
Aware Environment for Workshop with Game Storming .......................................................... 806
Noriyuki Kushiro, Tatsuya Ehira, and Rei Kaihara
BRPS: A Big Data Placement Strategy for Data Intensive Applications ...................................... 813
Lihui Liu, Junping Song, Haibo Wang, and Pin Lv
Data Exchange Platform to Fight Insurance Fraud on Blockchain ............................................ 821
Indranil Nath
Visualization Method Based on Contour Map ............................................................................. 826
Rui Yang and Yukio Ohsawa

OEDM — Optimization Based Techniques for Emerging Data Mining
MoCham: Robust Hierarchical Clustering Based on Multi-objective Optimization ...................... 831
Tomas Barton, Tomas Bruna, and Pavel Kordik
Structural Nonparallel Support Vector Machine Based on LSH for Large-Scale
Prediction .......................................................................... 839
Dandan Chen
Laplacian SVM for Learning from Label Proportions ............................................................... 847
Limeng Cui, Zhensong Chen, Fan Meng, and Yong Shi
Incorporate Hashing with Multi-view Learning  .................................................................................................................. 853  
*Jingjing Tang and Dewei Li*

**PDDM — Privacy and Discrimination in Data Mining**

Model-Based Approaches for Independence-Enhanced Recommendation  .................................................................................. 860  
*Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Issei Sato*

Classification Rule Mining Supported by Ontology for Discrimination Discovery  .............................................................................. 868  
*Binh Thanh Luong, Salvatore Ruggieri, and Franco Turini*

Faster Univariate Microaggregation for Power Law Distributions: 
k-Degree-Anonymity for Big Graphs  ............................................................................................................................................. 876  
*Julian Salas*

Scalable Privacy-Preserving Linking of Multiple Databases Using Counting Bloom Filters  ......................................................................................................................... 882  
*Dinusha Vatsalan, Peter Christen, and Erhard Rahm*

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

Research on Sentiment Analysis: The First Decade  ..................................................................................................................... 890  
*Oskar Ahlgren*

Mining the Opinionated Web: Classification and Detection of Aspect Contexts for Aspect Based Sentiment Analysis  .......................................................................................................................... 900  
*Oscar Araque, Ganggao Zhu, Manuel Garcia-Amado, and Carlos A. Iglesias*

What Drives Consumer Choices? Mining Aspects and Opinions on Large Scale Review Data Using Distributed Representation of Words  .................................................................................................................... 908  
*Kasturi Bhattacharjee and Linda Petzold*

Bayesian Deep Convolution Belief Networks for Subjectivity Detection  .............................................................................................. 916  
*Iti Chaturvedi, Erik Cambria, Soujanya Poria, and Rajiv Bajpai*

Sentiment Lexica from Paired Comparisons  ........................................................................................................................................... 924  
*Christoph Dalitz and Katrin E. Bednarek*

Seasonal Fluctuations in Collective Mood Revealed by Wikipedia Searches and Twitter Posts  ....................................................................................................................... 931  
*Fabon Dzogang, Thomas Lansdall-Welfare, and Nello Cristianini*

The Truth and Nothing But the Truth: Multimodal Analysis for Deception Detection  ........................................................................... 938  
*Mimansa Jaiswal, Sairam Tabibu, and Rajiv Bajpai*

Scalable and Real-Time Sentiment Analysis of Twitter Data  ................................................................................................................... 944  
*Maria Karanasou, Anetta Ampla, Christos Doulkeridis, and Maria Halkidi*

Multi-sentiment Modeling with Scalable Systematic Labeled Data Generation via Word2Vec Clustering  .......................................................................................................................... 952  
*Dhruv Mayank, Kanchana Padmanabhan, and Koushik Pal*
SmartVideoRanking: Video Search by Mining Emotions from Time-Synchronized Comments ................................................................. 960

Kosetsu Tsukuda, Hamasaki Masahiro, and Masataka Goto

Modeling Satire in English Text for Automatic Detection .................................................................................................................. 970

Aishwarya N. Reganti, Tushar Maheshwari, Upendra Kumar, Amitava Das, and Rajiv Bajpai

Lexicon Knowledge Extraction with Sentiment Polarity Computation .......................................................................................... 978

Zhaoxia Wang, Victor Joo Chuan Tong, Pingcheng Ruan, and Fang Li

SERecSys — Semantics-Enabled Recommender Systems

Exploiting a Determinant-Based Metric to Evaluate a Word-Embeddings Matrix of Items ................................................................. 984

Ludovico Boratto, Salvatore Carta, Gianni Fenu, and Roberto Saia

Semantic Enabled Recommender System for Micro-Blog Users ........................................................................................................ 992

Stefano Faralli, Giorgia Di Tommaso, and Paola Velardi

SemStim: Exploiting Knowledge Graphs for Cross-Domain Recommendation .................................................................................... 999

Benjamin Heitmann and Conor Hayes

Context-Specific Recommendation System for Predicting Similar PubMed Articles ............................................................... 1007

Shahin Mohammadi, Sudhir Kylasa, Giorgos Kollias, and Ananth Grama

SeBPR: Semantics Enhanced Bayesian Personalized Ranking with Comparable Item Pairs ............................................................... 1015

Penghua Yu and Lanfen Lin

SOMERIS — Social Media and Risk

Learning from User Workflows for the Characterization and Prediction of Software Crashes ................................................................. 1023

Chloé Adam, Antoine Aliotti, and Paul-Henry Cournède

Microblog Sentiment Topic Model ................................................................................................................................................. 1031

Aman Ahuja, Wei Wei, and Kathleen M. Carley

Graph-Based Term Weighting Scheme for Topic Modeling ............................................................................................................. 1039

Giannis Bekoulis and François Rousseau

Ensemble of Heterogeneous Classifiers for Improving Automated Tweet Classification .................................................. 1045

Renhao Cui, Gagan Agrawal, Rajiv Ramnath, and Vinh Khuc

Identifying Warning Behaviors of Violent Lone Offenders in Written Communication .................................................. 1053

Lisa Kaati, Amendra Shrestha, and Tony Sardella

Investors Attention and the Effects on Stock Market: An Empirical Study Based on Stock Forum .................................................. 1061

Wen Long, Lijing Guan, and Lingxiao Cui
Time-Based Ensembles for Prediction of Rare Events in News Stream .................................1066
    Nuno Moniz, Luis Torgo, and Magdalini Eirinaki

Fraud Detection in Voice-Based Identity Authentication Applications and Services ...............1074
    Saeid Safavi, Hock Gan, Iosif Mporas, and Reza Sotudeh

**SSTDM — Spatial and Spatiotemporal Data Mining**

Discovering Spatial Regions of High Correlation ....................................................................1082
    Prerna Agarwal, Richa Verma, and Venkata M. V. Gunturi

A Graph-Based Approach to Spatiotemporal Event Sequence Mining ....................................1090
    Berkay Aydin and Rafal A. Angryk

Hierarchical Aggregation Approach for Distributed Clustering of Spatial Datasets ..................1098
    Malika Bendechache, Nhien-An Le-Khac, and M-Tahar Kechadi

Markov Switching Copula Models for Longitudinal Data .......................................................1104
    Alfredo Cuesta-Infante and Kalyan Veeramachaneni

Location Recommendations for New Businesses Using Check-in Data ....................................1110
    Bahaeddin Eravci, Neslihan Bulut, Cagri Etemoglu, and Hakan Ferhatosmanoğlu

Generating Time-Varying Road Network Data Using Sparse Trajectories .............................1118
    Elif Eser, Furkan Kocayusufoglu, Bahaeddin Eravci, Hakan Ferhatosmanoğlu, and Josep L. Larriba-Pey

A Pattern Growth-Based Approach for Mining Spatiotemporal Co-occurrence Patterns ..........1125
    Shah Muhammad Hamdi, Berkay Aydin, and Rafal A. Angryk

Correlating Social Interconnections of Users with Spatio-Temporal Check-Ins
    Behavior ........................................................................................................................................1133
        Sonia Khetarpaul, S. K. Gupta, and L. Venkata Subramaniam

Flexible Similarity Search for Enriched Trajectories ................................................................1139
    Hideaki Ohashi, Toshiyuki Shimizu, and Masatoshi Yoshikawa

TrendTracker: Modelling the Motion of Trends in Space and Time ........................................1145
    Klaus Arthur Schmid, Christian Frey, Fengchao Peng, Michael Weiler, Andreas Züfle, Lei Chen, and Matthias Renz

A Genetic Algorithm to Discover Flexible Motifs with Support ...............................................1153
    Joan Serrá, Aleksandar Matic, Josep Lluis Arcos, and Alexandros Karatzoglou

Vehicle Classification from Low Frequency GPS Data ..............................................................1159
    Matteo Simoncini, Francesco Sambo, Leonardo Taecari, Luca Bravi, Samuele Salti, and Alessandro Lori

Hierarchical Temporal Memory Method for Time-Series-Based Anomaly Detection ...............1167
    Jia Wu, Weiru Zeng, Zhe Chen, and Xue-Fei Tang

Spatio-Temporal Site Recommendation ....................................................................................1173
    Guolei Yang and Andreas Züfle
PhD Forum

Sacrificing Overall Classification Quality to Improve Classification Accuracy of Well-Sought Classes ..........................................................1179
   Kevin M. Amaral, Ping Chen, Wei Ding, and Rajani Sadasivam

Towards Visualizing Hidden Structures ............................................................................................................................................1183
   Remy Dautriche, Alexandre Termier, Renaud Blanch, and Miguel Santana

Regression Techniques for Modelling Conception in Seasonally Calving Dairy Cows .................................................................1191
   Caroline Fenlon, Luke O’Grady, Michael Doherty, Stephen Butler, Laurence Shalloo, and John Dunnion

A Semi-Supervised Ensemble Approach for Multi-label Learning ........................................................................................................1197
   Ouadie Gharroudi, Haytham Elghazel, and Alexandre Aussem

A Novel Bayesian Ensemble Pruning Method ..........................................................................................................................1205
   Zhengshen Jiang, Hongzhi Liu, Bin Fu, and Zhonghai Wu

Improving the Prediction Cost of Drift Handling Algorithms by Abstaining .................................................................1213
   Pierre-Xavier Loeffel, Vincent Lemaire, Christophe Marsala, and Marcin Detyniecki

Infer Mobility Patterns and Social Dynamics for Modelling Human Behaviour .................................................................1223
   Luca Luceri

Subspace Clustering Ensembles through Tensor Decomposition ........................................................................................................1225
   Dominik Mautz, Christian Böhm, and Claudia Plant

Applying Deep Learning to Stereotypical Motor Movement Detection in Autism Spectrum Disorders ........................................................1235
   Nastaran Mohammadian Rad and Cesare Furlanello

Similarity Tree Pruning: A Novel Dynamic Ensemble Selection Approach ........................................................................................................1243
   Anil Narassiguin, Haytham Elghazel, and Alex Aussem

Clustering with the Levy Walk: “Hunting” for Clusters ..................................................................................................................1251
   Benjamin Schelling and Claudia Plant

Centrality-Based Approach for Supervised Term Weighting ........................................................................................................1261
   Niloofer Shanavas, Hui Wang, Zhiwei Lin, and Glenn Hawe

A Novel Pre-Classification Based kNN Algorithm ..................................................................................................................1269
   Huahua Xie, Dong Liang, Zhaojing Zhang, Hao Jin, Chen Lu, and Yi Lin

Demos

A Formation Energy Predictor for Crystalline Materials Using Ensemble Data Mining .................................................................1276
   Ankit Agrawal, Bryce Meredig, Chris Wolverton, and Alok Choudhary

Five Year Life Expectancy Calculator for Older Adults ..................................................................................................................1280
   Ankit Agrawal, Jason Mathias, David Baker, and Alok Choudhary
Financial Data Analysis with PGMs Using AMIDST .................................................................1284
  Rafael Cabañas, Ana M. Martínez, Andrés R. Masegosa, Dario Ramos-López,
  Antonio Samerón, Thomas D. Nielsen, Helge Langseth, and Anders L. Madsen

TENET: A Machine Learning-Based System for Target Characterization in Signaling
Networks ........................................................................................................................................1288

PaTSI: Pattern Mining of Time Series of Satellite Images in Knime ........................................1292
  Maxime Collin, Frédéric Flouvart, and Nazha Selmaoui-Folcher

airVLC: An Application for Visualizing Wind-Sensitive Interpolation of Urban Air
Pollution Forecasts .........................................................................................................................1296
  Lidia Contreras-Ochando and Cèsar Ferri

Explode: An Extensible Platform for Differentially Private Data Analysis ................................1300
  Emir Esmerdag, Mehmet Emre Gursay, Ali Inan, and Yucel Saygin

Cognito: Automated Feature Engineering for Supervised Learning ...........................................1304
  Udayan Khurana, Deepak Turaga, Horst Samulowitz, and Srinivasan Parthasarathy

Risk Mining: Company-Risk Identification from Unstructured Sources ......................................1308
  Timothy Nugent and Jochen L. Leidner

MobInsight: Understanding Urban Mobility with Crowd-Powered Neighborhood
Characterizations ............................................................................................................................1312
  Souneil Park, Marc Bourqui, and Enrique Frias-Martinez

Graph Mining for Complex Data Analytics ...................................................................................1316
  André Petermann, Martin Junghanns, Stephan Kemper, Kevin Gómez,
  Niklas Teichmann, and Erhard Rahm

FacetsViewer: A Tool for Multi-faceted Decomposition of Complex Networks ............................1320
  Boon-Siew Seah, Sourav S. Bhowmick, Huey Eng Chua, Mengxuan Chen,
  and C. Forbes Dewey Jr.

SpreadViz: Analytics and Visualization of Spreading Processes in Social Networks .....................1324
  Konstantinos Skianis, Maria-Evgenia G. Rossi, Fragkiskos D. Malliaros,
  and Michalis Vazirgiannis

DFA-G: A Unified Programming Model for Vertex-Centric Parallel Graph Processing ..................1328
  Bo Suo, Jing Su, Qun Chen, Zhanhuai Li, and Wei Pan

Informed Design Platform: Interpreting “Big Data” to Adaptive Place Designs .............................1332
  Linlin You and Bige Tunçer

Author Index