# Table of Contents

Welcome Message from the Conference Chairs.................................................................xx
Message from the Workshop Co-chairs...........................................................................xxii
Workshop Program Committee........................................................................................xxiii
Preface to the Fourth International Workshop on Mining
Multiple Information Sources......................................................................................xxviii
Preface to the Second Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes, and Impacts............................................................................xxix
Preface to the International Workshop on Knowledge Discovery Using Cloud and Distributed Computing Platforms.....................................................................................................xxx
Preface to the International Workshop on Spatial and Spatio-temporal Data Mining.........................................................................................................................xxxi
Preface to Social Interactions Analysis and Services Providers Workshop............................xxxii
Introduction to the Second IEEE International Workshop on Privacy Aspects of Data Mining.................................................................................................................xxxiv
Preface to the Third International Workshop on Semantic Aspects in Data Mining.........................................................................................................................xxxv
Preface to Biological Data Mining and its Applications in Healthcare...................................xxxvi
Preface to the Fifth International Workshop on Chance Discovery........................................xxxvii
Introduction to Reliability Issues in Knowledge Discovery..................................................xxxviii
Introduction to the Fourth International Workshop on Domain Driven Data Mining.....................xl
## Mining Multiple Information Sources—MMIS 2010

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>W-LEACH: Weighted Low Energy Adaptive Clustering Hierarchy Aggregation for Data Streams in Wireless Sensor Networks</td>
<td>1</td>
</tr>
<tr>
<td>Hanady M. Abdulsalam and Layla K. Kamel</td>
<td></td>
</tr>
<tr>
<td>Diffusion Maps: A Superior Semantic Method to Improve Similarity Join Performance</td>
<td>9</td>
</tr>
<tr>
<td>Bilal Hawashin, Farshad Fotouhi, and William Grosky</td>
<td></td>
</tr>
<tr>
<td>Mining Research Topics Evolving Over Time Using a Diachronic Multi-source Approach</td>
<td>17</td>
</tr>
<tr>
<td>Jean-Charles Lamirel, Ghada Safi, Navesh Priyankar, and Pascal Cuxac</td>
<td></td>
</tr>
<tr>
<td>PTCR-Miner: Progressive Temporal Class Rule Mining for Multivariate Temporal Data Classification</td>
<td>25</td>
</tr>
<tr>
<td>Chao-Hui Lee and Vincent S. Tseng</td>
<td></td>
</tr>
<tr>
<td>Differential Analysis on Deep Web Data Sources</td>
<td>33</td>
</tr>
<tr>
<td>Tantan Liu, Fan Wang, Jiedan Zhu, and Gagan Agrawal</td>
<td></td>
</tr>
<tr>
<td>Learning a Combination of Dissimilarities from a Set of Equivalence Constraints</td>
<td>41</td>
</tr>
<tr>
<td>Manuel Martin-Merino</td>
<td></td>
</tr>
<tr>
<td>A Triple-Random Ensemble Classification Method for Mining Multi-label Data</td>
<td>49</td>
</tr>
<tr>
<td>Gulisong Nasierding, Abbas Z. Kouzani, and Grigoris Tsoumakas</td>
<td></td>
</tr>
<tr>
<td>Learning Document Labels from Enriched Click Graphs</td>
<td>57</td>
</tr>
<tr>
<td>Lan Nie, Zhigang Hua, Xiaofeng He, and Scott Gaffney</td>
<td></td>
</tr>
<tr>
<td>Multiple Feature-Based Classifier and Its Application to Image Classification</td>
<td>65</td>
</tr>
<tr>
<td>Dong-Chul Park</td>
<td></td>
</tr>
<tr>
<td>CUBS: Multivariate Sequence Classification Using Bounded Z-score with Sampling</td>
<td>72</td>
</tr>
<tr>
<td>Ariella Richardson, Gal Kaminka, and Sarit Kraus</td>
<td></td>
</tr>
<tr>
<td>Event Data Mining and Classification from Multiple Streaming Sources</td>
<td>80</td>
</tr>
<tr>
<td>Ashit Talukder</td>
<td></td>
</tr>
<tr>
<td>Distributed Classification on Peers with Variable Data Spaces and Distributions</td>
<td>88</td>
</tr>
<tr>
<td>Quach Vinh Thanh, Vivekanand Gopalkrishnan, and Hock Hee Ang</td>
<td></td>
</tr>
</tbody>
</table>
**Knowledge Discovery from Climate Data: Prediction, Extremes, and Impacts—CLIMKD 2010**

Detecting Climate Change in Multivariate Time Series Data by Novel Clustering and Cluster Tracing Techniques .................................................................96

*Hardy Kremer, Stephan Günnemann, and Thomas Seidl*

A New Data Mining Model for Hurricane Intensity Prediction ..........................................................................................................................98

*Yu Su, Sudheer Chelluboina, Michael Hahsler, and Margaret H. Dunham*

Mining Spatial Co-occurrence of Drought Events from Climate Data of India ..........................................................106

*Kollukuduru Sravanthi and K.S. Rajan*

Data Analysis in Los Angeles Long Beach with Seasonal Time Series Model ..........................................................113

*Weiqiang Wang and Zhendong Niu*

---

**Knowledge Discovery Using Cloud and Distributed Computing Platforms—KDCloud 2010**

On Finding Similar Items in a Stream of Transactions ..........................................................121

*Andrea Campagna and Rasmus Pagh*

Minimum Spanning Tree Based Classification Model for Massive Data with MapReduce Implementation ..................................................................................................................129

*Jin Chang, Jun Luo, Joshua Zhexue Huang, Shengzhong Feng, and Jianping Fan*

Distributed, Scalable Clustering for Detecting Halos in Terascale Astronomy Datasets ........................................................138

*Srivatsava Daruru, Sankari Dhandapani, Gunjan Gupta, Ilian Iliev, Weijia Xu, Paul Navratil, Nena Marin, and Joydeep Ghosh*

Challenges in Scheduling Aggregation in Cyberphysical Information Processing Systems ..........................................................148

*James L. Horey*

Parallel User Profiling Based on Folksonomy for Large Scaled Recommender Systems: An Implementation of Cascading MapReduce ..........................................................154

*Huizhi Liang, Jim Hogan, and Yue Xu*

dSimpleGraph: A Novel Distributed Clustering Algorithm for Exploring Very Large Scale Unknown Data Sets ..........................................................162

*Li Lu, Yunhong Gu, and Robert Grossman*

S4: Distributed Stream Computing Platform ..........................................................170

*Leonardo Neumeyer, Bruce Robbins, Anish Nair, and Anand Kesari*

Parallel EM-Clustering: Fast Convergence by Asynchronous Model Updates ..........................................................178

*Claudia Plant and Christian Böhnm*

Parallelizing an Information Theoretic Co-clustering Algorithm Using a Cloud Middleware ..........................................................186

*Venkatram Ramanathan, Wenjing Ma, Vignesh T. Ravi, Tantan Liu, and Gagan Agrawal*
Spatial and Spatio-temporal Data Mining—SSTDM 2010

Putting Pixels in Place: A Storage Layout Language for Scientific Data ................................................................. 194
  Peter Baumann, Shams Feyzabadi, and Constantin Jucovschi

Using Time Series Segmentation for Deriving Vegetation Phenology Indices from MODIS NDVI Data ................................................................. 202
  Varun Chandola, Dafeng Hui, Lianhong Gu, Budhendra Bhaduri, and Ranga Raju Vatsavai

Comparing Vessel Trajectories Using Geographical Domain Knowledge and Alignments ................................................................. 209
  Gerben K.D. de Vries, Willem Robert van Hage, and Maarten van Someren

Learning Maximum Lag for Grouped Graphical Granger Models ............................................................................. 217
  Amit Dhurandhar

Using Self-Organizing Map and Heuristics to Identify Small Statistical Areas Based on Household Socio-Economic Indicators in Turkey’s Address Based Population

Register System ................................................................................................................................................................. 225
  Sebnem Düzgün and Seyma Özcan Yavuzoglu

Computing Popular Places Using Graphics Processors .................................................................................. 233
  Marta Fort, J. Antoni Sellarès, and Nacho Valladares

Sample Bias due to Missing Data in Mobility Surveys .................................................................................. 241
  Dirk Hecker, Hendrik Stange, Christine Körner, and Michael May

Spatio-Temporal Symbolization of Multidimensional Time Series ............................................................................. 249
  Shohei Hidaka and Chen Yu

Curvature Maxima-based Trajectories Mining ............................................................................................................ 257
  Shoji Hirano and Shusaku Tsumoto

A K-Main Routes Approach to Spatial Network Activity Summarization: A Summary of Results ............................................................................. 265
  Dev Oliver, Abdussalam Bannur, James M. Kang, Shashi Shekhar, and Renee Bousselaire

  Ranga Raju Vatsavai, Anil Cheriyadat, and Shaun Gleason

Measuring Similarity for Multidimensional Sequences .................................................................................. 281
  Hui Wang, Zhiwei Lin, Sally McClean, and Jun Liu

Social Interactions Analysis and Services Providers—SIASP 2010

Towards Community Discovery in Signed Collaborative Interaction Networks ................................................................. 288
  Petko Bogdanov, Nicholas D. Larusso, and Ambuj Singh

Catching a Viral Video .................................................................................................................................................. 296
  Tom Broxton, Yannet Interian, Jon Vaver, and Mirjam Wattenhofer
Improving Matching Process in Social Network .......................................................... 305
  Lin Chen, Richi Nayak, and Yue Xu

Extracting Representative Tags for Flickr Users ...................................................... 312
  Xian Chen and Hyoseop Shin

Detecting and Tracking Community Dynamics in Evolutionary Networks ................. 318
  Zhengzhang Chen, Kevin A. Wilson, Ye Jin, William Hendrix, and Nagiza F. Samatova

The Meme Ranking Problem: Maximizing Microblogging Virality ........................... 328
  Dino Ienco, Francesco Bonchi, and Carlos Castillo

Modeling and Comparing the Influence of Neighbors on the Behavior of Users in Social and Similarity Networks .......................................................... 336
  Mohsen Jamali and Martin Ester

Modeling the Temporal Dynamics of Social Rating Networks Using Bidirectional Effects of Social Relations and Rating Patterns ............................................. 344
  Mohsen Jamali, Gholamreza Haffari, and Martin Ester

A Social Matching System for an Online Dating Network: A Preliminary Study ............ 352
  Richi Nayak, Meng Zhang, and Lin Chen

RnR: Extracting Rationale from Online Reviews and Ratings .................................... 358
  Dwi AP Rahayu, Shonali Krishnaswamy, Oshadi Alahakoon, and Cyril Labbe

Bridging Folksonomies and Domain Ontologies: Getting Out Non-taxonomic Relations ........................................................................................................... 369
  Chiraz Trabelsi, Aicha Ben Jrad, and Sadok Ben Yahia

Toward Finding Hidden Communities Based on User Profile .................................... 380
  Tetsuya Yoshida

Visual Analytics and Knowledge Discovery—VAKD 2010

A Visual Analytics Tool for Analysing Microarray Data ........................................... 388
  Martin Amélie, Monique Noirhomme-Fraiture, Quang Vinh Nguyen, and Simeon Simoff

Analysing the Large-Scale, Time-Oriented, VAST 2010 MC2 Dataset with PRISMA .......... 396
  Rafael Guimarães, Nicolas Carneiro, Anderson Marques, Bianchi Meiguins, and Aruanda Simões Meiguins

Evolving Ensemble-Clustering to a Feedback-Driven Process ..................................... 401
  Martin Hahmann, Dirk Habich, and Wolfgang Lehner

Visually Controllable Data Mining Methods ......................................................... 409
  Kai Puolamäki, Panagiotis Papapetrou, and Jeffrey Lijffijt

User-Based Active Learning ..................................................................................... 418
  Christin Seifert and Michael Granitzer

SIMPLE: Interactive Analytics on Patent Data ......................................................... 426
  Scott Spangler, Ying Chen, Jeffrey Kreulen, Stephen Boyer, Thomas Griffin, Alfredo Alba, Linda Kato, Ana Lelescu, and Su Yan
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Dimensional Multimodal Distribution Embedding</td>
<td>434</td>
</tr>
<tr>
<td>Eniko Szekely, Eric Bruno, and Stephane Marchand-Maillet</td>
<td></td>
</tr>
<tr>
<td>BODY—Buckets of Disease Symptoms for Disease Outbreak Analysis</td>
<td>442</td>
</tr>
<tr>
<td>Hanisha Veeramachaneni, Soujanya Vadapalli, and Kamalakar Karlapalem</td>
<td></td>
</tr>
<tr>
<td>Privacy Aspects of Data Mining—PADM 2010</td>
<td></td>
</tr>
<tr>
<td>Language-Based Enforcement of Access Constraints on Geoscientific Image Data</td>
<td>450</td>
</tr>
<tr>
<td>Peter Baumann</td>
<td></td>
</tr>
<tr>
<td>Identifying and Preventing Data Leakage in Multi-relational Classification</td>
<td>458</td>
</tr>
<tr>
<td>Hongyu Guo, Herna L. Viktor, and Eric Paquet</td>
<td></td>
</tr>
<tr>
<td>On Attribute Disclosure in Randomization Based Privacy Preserving Data Publishing</td>
<td>466</td>
</tr>
<tr>
<td>Ling Guo, Xiaowei Ying, and Xintao Wu</td>
<td></td>
</tr>
<tr>
<td>Privacy Violations Using Microtargeted Ads: A Case Study</td>
<td>474</td>
</tr>
<tr>
<td>Aleksandra Korolova</td>
<td></td>
</tr>
<tr>
<td>Frequent Closed Itemset Mining with Privacy Preserving for Distributed Databases</td>
<td>483</td>
</tr>
<tr>
<td>Shin-ya Kuno, Koichiro Doi, and Akihiro Yamamoto</td>
<td></td>
</tr>
<tr>
<td>Anonymizing Graphs Against Weight-Based Attacks</td>
<td>491</td>
</tr>
<tr>
<td>Yidong Li and Hong Shen</td>
<td></td>
</tr>
<tr>
<td>A Privacy Preserving Framework for Gaussian Mixture Models</td>
<td>499</td>
</tr>
<tr>
<td>Madhusudana Shashanka</td>
<td></td>
</tr>
<tr>
<td>Identifying Similar Neighborhood Structures in Private Social Networks</td>
<td>507</td>
</tr>
<tr>
<td>Lisa Singh and Clare Schramm</td>
<td></td>
</tr>
<tr>
<td>Semantic Aspects in Data Mining—SADM 2010</td>
<td></td>
</tr>
<tr>
<td>Geospatial Schema Matching with High-Quality Cluster Assurance and Location</td>
<td>517</td>
</tr>
<tr>
<td>Latifur Khan, Jeffrey Partyka, Satyen Abrol, and Bhavani Thuraisingham</td>
<td></td>
</tr>
<tr>
<td>Semantic Content Filtering with Wikipedia and Ontologies</td>
<td>518</td>
</tr>
<tr>
<td>Pekka Malo, Pyry Siitari, Oskar Ahlgren, Jyrki Wallenius, and Pekka Korhonen</td>
<td></td>
</tr>
<tr>
<td>From Pattern Discovery to Pattern Interpretation in Movement Data</td>
<td>527</td>
</tr>
<tr>
<td>Rebecca Ong, Monica Wachowicz, Mirco Nanni, and Chiara Renso</td>
<td></td>
</tr>
<tr>
<td>Incorporating Multi-partite Networks and Expertise to Construct Related-Term Graphs</td>
<td>535</td>
</tr>
<tr>
<td>Jyh-Ren Shieh, Ching-Yung Lin, Shun-Xuan Wang, Yung-Huan Hsieh, and Ja-Ling Wu</td>
<td></td>
</tr>
<tr>
<td>Analysis of Collaborative Writing Processes Using Hidden Markov Models and Semantic Heuristics</td>
<td>543</td>
</tr>
<tr>
<td>Vilaythong Southavilay, Kalina Yacef, and Rafael A. Calvo</td>
<td></td>
</tr>
<tr>
<td>Using Ontology and Sequence Information for Extracting Behavior Patterns from Web Navigation Logs</td>
<td>549</td>
</tr>
<tr>
<td>Hakan Yilmaz and Pinar Senkul</td>
<td></td>
</tr>
</tbody>
</table>
Applications of Data Mining and Modeling in Government and Industry—ADMMGI 2010

Variable Productivity Adjustment Estimation for Function Point Project Delivery ........................................ 557
    Saeed Bagheri, Krishna Ratakonda, and Rakesh Mohan

Graphics Classification for Enterprise Knowledge Management ................................................................. 562
    Divna Djordjevic and Rayid Ghani

Efficient Rule Generation for Dominant Class Problems on LARM ......................................................... 570
    JuiHsi Fu and SingLing Lee

Application of Data Mining for Anti-money Laundering Detection: A Case Study ........................................ 577
    Nhien An Le Khac and M-Tahar Kechadi

Empirical Analysis: News Impact on Stock Prices Based on News Density ................................................. 585
    Xiaodong Li, Xiaotie Deng, Feng Wang, and Keren Dong

Sales Forecasting of IT Products Using a Hybrid MARS and SVR Model ................................................... 593
    Chi-Jie Lu, Tian-Shyug Lee, and Chia-Mei Lian

Unsupervised DRG Upcoding Detection in Healthcare Databases ......................................................... 600
    Wei Luo and Marcus Gallagher

Mining Heterogeneous ADS-B Data Sets for Probabilistic Models of Pilot Behavior .................................... 606
    Ron Marsh and Kirk Ogaard

Detecting Non-compliant Consumers in Spatio-Temporal Health Data: A Case Study from Medicare Australia ................................................................. 613
    K.S. Ng, Y. Shan, D.W. Murray, A. Sutinen, B. Schwarz, D. Jeacocke, and J. Farrugia

Visual Predictions of Currency Crises Using Self-Organizing Maps ....................................................... 623
    Peter Sarlin and Dorina Marghescu

Gaussian Processes for Dispatching Rule Selection in Production Scheduling: Comparison of Learning Techniques ................................................................. 631
    Bernd Scholz-Reiter, Jens Heger, and Torsten Hildebrandt

Using SOM-Ward Clustering and Predictive Analytics for Conducting Customer Segmentation .................. 639
    Zhiyuan Yao, Tomas Eklund, and Barbro Back

Quantum Path Integral Inspired Query Sequence Suggestion for User Search Task Simplification ................. 647
    Baojun Yue, Jun Yan, Heng Liang, Ning Liu, Lei Ji, Fengshan Bai, and Zheng Chen

Less Effort, More Outcomes: Optimising Debt Recovery with Decision Trees ........................................ 655
    Yanchang Zhao, Hans Bohlscheid, Shanshan Wu, and Longbing Cao
Biological Data Mining and Its Applications in Healthcare—BioDM 2010

System Biology Approach for Elucidating the Relationship Between Indonesian Herbal Plants and the Efficacy of Jamu ......................................................................................................................661

The Multiple Alignment Algorithm for Metabolic Pathways without Abstraction ......................................................669
  Wenbin Chen, Andrea M. Rocha, William Hendrix, Matthew Schmidt, and Nagiza F. Samatova

Fixed-Parameter Tractable Combinatorial Algorithms for Metabolic Networks Alignments ..................................................................................................................................................................679
  Qiong Cheng, Jinpeng Wei, Alexander Zelikovsky, and Mitsunori Oghara

Efficient Alignments of Metabolic Networks with Bounded Treewidth ........................................................................687
  Qiong Cheng, Piotr Berman, Rob Harrison, and Alex Zelikovsky

Curve Profiling Feature: Novel Compact Representation for Drosophila Embryonic Gene Expression Pattern Mining ..............................................................................................................................................695
  Chunsheng Fang, Minlu Zhang, Anca L. Ralescu, and Jason L. Lu

A Sequence Data Mining Protocol to Identify Best Representative Sequence for Protein Domain Families ........................................................................................................................................................................703
  V.S. Gowri, Khader Shameer, Chilamakuri Chandra Sekhar Reddy, Prashant Shingate, and Ramanathan Sowdhamini

Parametric Templates: A New Enzyme Active-Site Prediction Algorithm ........................................................................711
  Tsuyoshi Kato, Kazuhiro Suwa, and Nozomi Nagano

Overcoming the Curse of Dimensionality in a Statistical Geometry Based Computational Protein Mutagenesis ..................................................................................................................719
  Majid Masso

Interaction as an Interestingness Measure ..................................................................................................................726
  Martin McGrane and Simon K. Poon

Is Topotecan Effective at Killing Cancer Cells? ..............................................................................................................732
  Ricardo Santiago-Mozos, Imtiaz A. Khan, and Michael G. Madden

Cost-Sensitive Feature Selection Based on the Set Covering Machine ........................................................................740
  Raúl Santos-Rodríguez and Dario García-García

Combining Time Series Similarity with Density-Based Clustering to Identify Fiber Bundles in the Human Brain ..................................................................................................................747
  Junming Shao, Klaus Hahn, Qinli Yang, Christian Böhm, Afra Wohlschläger, Nicholas Myers, and Claudia Plant

Evaluation of Protein Backbone Alphabets: Using Predicted Local Structure for Fold Recognition .................................................................................................................................................755
  Kyong Jin Shim
Fast Anomaly Detection in Dynamic Clinical Datasets Using Near-Optimal Hashing with Concentric Expansions
Zeeshan Syed and Ilan Rubinfeld

An Empirical Comparison of Platt Calibration and Inductive Confidence Machines for Predictions in Drug Discovery
Nikil Wale

FDCluster: Mining Frequent Closed Discriminative Bicluster without Candidate Maintenance in Multiple Microarray Datasets
Miao Wang, Xueqin Shang, Shaohua Zhang, and Zhanhuai Li

An Integrative Scoring Approach to Identify Transcriptional Regulations Controlling Lung Surfactant Homeostasis
Minlu Zhang, Chunsheng Fang, Yan Xu, Raj K. Bhatnagar, and Long J. Lu

International Workshop on Chance Discovery—IWCD 2010

Curation in Chance Discovery
Akinori Abe

Text Categorization with Considering Temporal Patterns of Term Usages
Hidenao Abe and Shusaku Tsumoto

Paired Evaluators Method to Track Concept Drift: An Application for Hedge Funds Operations
Masabumi Furuhata, Takanobu Mizuta, and Jihei So

Innovators Marketplace as Game on the Table versus Board on the Web
Yukio Ohsawa, Kensuke Okamoto, Yuma Takahashi, and Yoko Nishihara

Finding Rare Patterns with Weak Correlation Constraint
Yoshiaki Okubo, Makoto Haraguchi, and Takeshi Nakajima

Temporal Logic $\mathcal{TL} ^{\mathcal{Z} \cup \{DU\}}$ Modeling Local and Global Discovery with Logical Uncertainty
Vladimir V. Rybakov

Automated Empirical Selection of Rule Induction Methods Based on Recursive Iteration of Resampling Methods and Multiple Testing
Shusaku Tsumoto and Shoji Hirano

Change with Delayed Labeling: When is it Detectable?
Indrė Žliobaitė

Reliability Issues in Knowledge Discovery—RIKD 2010

Towards a Reliable Framework of Uncertainty-Based Group Decision Support System
Junyi Chai and James N.K. Liu
A Study on the Accuracy of Frequency Measures and Its Impact on Knowledge Discovery in Single Sequences .................................................................................................................................. 859
  
  Min Gan and Honghua Dai

Co-occurrence Based Predictors for Estimating Query Difficulty ................................................................. 867
  
  Hazra Imran and Aditi Sharan

Batch Mode Sparse Active Learning .................................................................................................................. 875
  
  Lixin Shi and Yuhang Zhao

Towards the Detection of Potential Contradictions in Fuzzy Ontology Using a High Level Net Approach Integrated with Uncertainty Inference .................................................................................. 883
  
  Ke Wang, James N.K. Liu, and Wei-Min Ma

Learning Robust Bayesian Network Classifiers in the Space of Markov Equivalent Classes ................................................................. 891
  
  Zhongfeng Wang, Zhihai Wang, and Bin Fu

Learning Restricted Bayesian Network Classifiers with Mixed Non-i.i.d. Sampling .............................................. 899
  
  Zhongfeng Wang, Zhihai Wang, and Bin Fu

Semantic Image Retrieval Based on Multiple-Instance Learning ........................................................................ 905
  
  Min Yao, Wenshen Yi, Rong Zhu, and Ran Cheng

Domain Driven Data Mining—DDDM 2010

Link Prediction Across Multiple Social Networks ............................................................................................. 911
  
  Muhammad Aurangzeb Ahmad, Zoheb Borbora, Jaideep Srivastava, and Noshir Contractor

Controlling Consistency in Top-N Recommender Systems ............................................................................... 919
  
  Paolo Cremonesi and Roberto Turrin

Evaluating Association Rules by Quantitative Pairwise Property Comparisons .................................................. 927
  
  Elnaz Delpisheh and John Z. Zhang

Trading Tests of Long-Term Market Forecast by Text Mining.............................................................................. 935
  
  Kiyoshi Izumi, Takashi Goto, and Tohgoroh Matsui

Network Anomaly Detection Using a Commute Distance Based Approach ....................................................... 943
  
  Nguyen Lu Dang Khoa, Tahereh Babaie, Sanjay Chawla, and Zainab Zaidi

Vote-Based LELE for Positive and Unlabeled Textual Data Streams .................................................................. 951
  
  Bo Liu, Yanshan Xiao, Longbing Cao, and Philip S. Yu

Domain-Driven Data Mining for IT Infrastructure Support ............................................................................... 959
  
  Girish Keshav Palshikar, Harrick M. Vin, Mohammed Mudassar, and Maitreya Natu

Insights from Applying Sequential Pattern Mining to E-commerce Click Stream Data ..................................... 967
  
  Arthur Pitman and Markus Zanker

Medically Driven Data Mining Application: Recognition of Health Problems from Gait Patterns of Elderly ...................................................................................................................................... 976
  
  Bogdan Pogorelc and Matjaž Gams
Contextual Sequential Pattern Mining .................................................................981
    Julien Rabatel, Sandra Bringay, and Pascal Foncelet

Using Taxonomies to Perform Aggregated Querying over Imprecise Data ........................................989
    Atanu Roy, Chandrima Sarkar, and Rafal A. Angryk

Sequence Alignment Based Analysis of Player Behavior in Massively Multiplayer
Online Role-Playing Games (MMORPGs) ..............................................................997
    Kyong Jin Shim and Jaideep Srivastava

Agent Assignment for Process Management: Agent Performance Evaluation
Framework ...............................................................................................................1005
    Ramzan Talib, Bernhard Volz, and Stefan Jablonski

The Effectiveness of a New Negative Correlation Learning Algorithm
for Classification Ensembles ......................................................................................1013
    Shuo Wang and Xin Yao

Learning with Actionable Attributes: Attention—Boundary Cases! ................................................1021
    Indre Žliobaitė and Mykola Pechenizkiy

Data Mining for Service—DMS 2010
Large-Scale Customized Models for Advertisers ......................................................1029
    Abraham Bagherjeiran, Andrew Hatch, Adwait Ratnaparkhi, and Rajesh Parekh

Knowledge File System—A Principled Approach to Personal Information
Management .............................................................................................................1037
    Kuiyu Chang, I. Wayan Tresna Perdana, Bramandia Ramadhana, Kailash Sethuraman,
    Truc Viet Le, and Neha Chachra

Automated Prompting in a Smart Home Environment .............................................1045
    Barnan Das, Chao Chen, Nairanjana Dasgupta, Diane J. Cook, and Adriyana M. Seelye

Contextual Segmentation: Using Context to Improve Behavior Predictive Models
in E-commerce ..........................................................................................................1053
    Maria Francesca Faraone, Michele Gorgoglione, and Cosimo Palmisano

Interesting Subset Discovery and Its Application on Service Processes ..................1061
    Maitreya Natu and Girish Keshav Palshikar

Discovering Temporal Features and Relations of Activity Patterns .........................1069
    Ehsan Nazerfard, Parisa Rashidi, and Diane J. Cook

Towards Data-Oriented Hospital Services: Data Mining-Based Hospital Management ........................................1076
    Shusaku Tsumoto, Shoji Hirano, and Yuko Tsumoto
International Workshop on Topic Feature Discovery and Opinion Mining—FDOM 2010

Infrequent Purchased Product Recommendation Making Based on User Behaviour and Opinions in E-commerce Sites ................................................................................................................1084
Noraswaliza Abdullah, Yue Xu, Shlomo Geva, and Jinghong Chen

The PARIS Algorithm for Determining Latent Topics ............................................................................................1092
Michal Aharon, Ira Cohen, Arik Itskovitch, Inbal Marhaim, and Ron Banner

A Block Mixture Model for Pattern Discovery in Preference Data ......................................................................1100
Nicola Barbieri, Massimo Guarascio, and Giuseppe Manco

Mining Arabic Business Reviews .............................................................................................................................1108
Mohamed Elhawary and Mohamed Elfeky

Sentence-Level and Document-Level Sentiment Mining for Arabic Texts ..............................................................1114
Noura Farra, Elie Challita, Rawad Abou Assi, and Hazem Hajj

High-Order Concept Associations Mining and Inferential Language Modeling for Online Review Spam Detection .........................................................................................................................1120
C.L. Lai, K.Q. Xu, Raymond Y.K. Lau, Yuefeng Li, and Dawei Song

Mining Users’ Opinions Based on Item Folksonomy and Taxonomy for Personalized Recommender Systems .................................................................................................................................1128
Huizhi Liang, Yue Xu, and Yuefeng Li

A Framework for Emotion Mining from Text in Online Social Networks .................................................................................................................................................................................................1136
Mohamed Yassine and Hazem Hajj

Augmenting Chinese Online Video Recommendations by Using Virtual Ratings Predicted by Review Sentiment Classification ........................................................................................................1143
Weishi Zhang, Guiguang Ding, Li Chen, and Chunping Li

Optimization Based Techniques for Emerging Data Mining Problems—OEDM 2010

ValuePick: Towards a Value-Oriented Dual-Goal Recommender System ..............................................................1151
Leman Akoglu and Christos Faloutsos

Efficient Dimensionality Reduction on Undersampled Problems through Incremental Discriminative Common Vectors ................................................................................................................1159
Francesc J. Ferri, Katerine Diaz-Chito, and Wladimiro Diaz-Villanueva

XML Documents Clustering Using Tensor Space Model—A Preliminary Study ..................................................1167
Sangetha Kutty, Richi Nayak, and Yuefeng Li

Ant Colony Optimization Algorithm Based on Immunity Vaccine and Dynamic Pheromone Updating ..........................................................................................................................................................1174
Wanjun Liu, Juan Zhang, and Junli Liu

Robust Low-Rank Subspace Segmentation with Semidefinite Guarantees ..........................................................1179
Yuzhao Ni, Ju Sun, Xiaotong Yuan, Shuicheng Yan, and Loong-Fah Cheong
Find Intelligent Knowledge by Second-Order Mining: Three Cases from China .................................................................1189

Guangli Nie, Lingling Zhang, Yuejin Zhang, Wei Deng, and Yong Shi

Semi-supervised PLSA for Document Clustering ........................................................................................................1196

Lingfeng Niu and Yong Shi

Cluster Cores and Modularity Maximization .............................................................................................................1204

Michael Ovelgönne and Andreas Geyer-Schulz

Reviewer Profiling Using Sparse Matrix Regression ..................................................................................................1214

Evangelos E. Papalexakis, Nicholas D. Sidiropoulos, and Minos N. Garofalakis

Distributed Flow Algorithms for Scalable Similarity Visualization ..............................................................................1220

Novi Quadrianto, Dale Schuurmans, and Alex J. Smola

Efficient Additive Models via the Generalized Lasso ........................................................................................................1228

Dimitri Semenovich, Nobuyuki Morioka, and Arcot Sowmya

A Comparison of Objective Functions in Network Community Detection ........................................................................1234

Chuan Shi, Yanan Cai, Philip S. Yu, Zhenyu Yan, and Bin Wu

Large-Scale Matrix Factorization Using MapReduce ..................................................................................................1242

Zhengguo Sun, Tao Li, and Naphali Rishe

Integrating Symmetric Nonnegative Matrix Factorization and Normalized Cut Spectral Clustering ........................................1249

Zhichen Xia and Chris Ding

Integer Programming for Multi-class Active Learning ......................................................................................................1257

Dragomir Yankov, Suju Rajan, and Adwait Ratnaparkhi

ALPOS: A Machine Learning Approach for Analyzing Microblogging Data ......................................................................1265

Dan Zhang, Yan Liu, Richard D. Lawrence, and Vijil Chenthamarakshan

A Novel Ranking Algorithm for Service Matching Based on Agent Association Graphs ....................................................1273

Hao Lan Zhang, Clement H.C. Leung, Gitesh K. Raikundalia, and Jing He

From Convex to Nonconvex: A Loss Function Analysis for Binary Classification .............................................................1281

Lei Zhao, Musa Mammadov, and John Yearwood

Subspace Distance-Based Sampling Method for SVM .......................................................................................................1289

Xiaofei Zhou and Yong Shi

Large-Scale Analytics for Complex Instrumented Systems—LACIS 2010

Sliding HyperLogLog: Estimating Cardinality in a Data Stream over a Sliding Window .................................................................1297

Yousra Chabchoub and Georges Hébrail

Using Adaptive Downsampling to Compare Time Series with Warping ..............................................................................1304

Chih-Chun Chia and Zeeshan Syed
Multi-step Time Series Prediction in Complex Instrumented Domains

Amit Dhurandhar

1312

dMaximalCliques: A Distributed Algorithm for Enumerating All Maximal Cliques and Maximal Clique Distribution

Li Lu, Yunhong Gu, and Robert Grossman

1320

EigenDiagnostics: Spotting Connection Patterns and Outliers in Large Graphs

Koji Maruhashi and Christos Faloutsos

1328

Enhancing Ubiquitous Systems through System Call Mining

Katharina Morik, Felix Jungermann, Nico Piatkowski, and Michael Engel

1338

Parallelized Boosting with Map-Reduce

Indranil Palit and Chandan K. Reddy

1346

Contest Papers

A Convex Combination of Models for Predicting Road Traffic

Carlos J. Gil Bellosta

1354

Predicting Travel Times with Context-Dependent Random Forests by Modeling Local and Aggregate Traffic Flow

Benjamin Hamner

1357

Predicting Future Traffic Congestion from Automated Traffic Recorder Readings with an Ensemble of Random Forests

Benjamin Hamner

1360

Ensemble-Based Method for Task 2: Predicting Traffic Jam

Jingrui He, Qing He, Grzegorz Swirszcz, Yiannis Kamarianakis, Rick Lawrence, Wei Shen, and Laura Wynter

1363

IEEE ICDM 2010 Contest: Traffic Prediction—Jams

Lukasz Romaszko

1366

Traffic Velocity Prediction Using GPS Data: IEEE ICDM Contest Task 3 Report

Wei Shen, Yiannis Kamarianakis, Laura Wynter, Jingrui He, Qing He, Rick Lawrence, and Grzegorz Swirszcz

1369


Marcin Wojnarski, Pawel Gora, Marcin Szczuka, Hung Son Nguyen, Joanna Swietlicka, and Demetris Zeinalipour

1372

Demo Papers

Meerkat: Community Mining with Dynamic Social Networks

Jiyang Chen, Justin Fagnan, Randy Goebel, Reihaneh Rabbany, Farzad Sangi, Mansoureh Takaffoli, Eric Verbeek, and Osmar Zaíane

1377
Clutter-Adaptive Visualization for Mobile Data Mining .................................................................1381

Brett Gillick, Hasnain AlTaiair, Shonali Krishnaswamy, Jonathan Liono, Nicholas Nicoloudis, Abhijat Sinha, Arkady Zaslavsky, and Mohamed Medhat Gaber

Data Prediction Competitions—Far More than Just a Bit of Fun .........................................................1385

Anthony Goldbloom

MCExplorer: Interactive Exploration of Multiple (Subspace) Clustering Solutions ..............................1387

Stephan Günnemann, Hardy Kremer, Ines Färber, and Thomas Seidl

Targeting Television Audiences Using Demographic Similarity ........................................................1391

Brendan Kitts, Liang Wei, Dyng Au, Stefanie Zlomek, Ryan Brooks, and Brian Burdick

Clustering Performance on Evolving Data Streams: Assessing Algorithms and Evaluation Measures within MOA .................................................................1400

Philipp Kranen, Hardy Kremer, Timm Jansen, Thomas Seidl, Albert Bifet, Geoff Holmes, and Bernhard Pfahringer

i-Analyst: An Agent-Based Distributed Data Mining Platform ..............................................................1404

Chayapol Moemeng, Xinhua Zhu, Longbing Cao, and Chen Jiahang

Demo 2009

Enhancing Document Exploration with OLAP ......................................................................................1407

Zhibo Chen, Carlos Garcia-Alvarado, and Carlos Ordonez

QueRIE: A Query Recommender System Supporting Interactive Database Exploration ......................1411

Sarika Mittal, Jothi Swarubini Vindhiya Varman, Gloria Chatzopoulou, Magdalini Eirinaki, and Neoklis Polyzotis

uRule: A Rule-Based Classification System for Uncertain Data .................................................................1415

Biao Qin, Yuni Xia, Rakshesh Sathyesh, Sunil Prabhukar, and Yicheng Tu

Adaptive Multimedia Mining on Distributed Stream Processing Systems .................................................1419

Deepak S. Turaga, Hyunggon Park, Rong Yan, and Olivier Verscheure

Author Index