

2010 IEEE International Conference on Data Mining Workshops

(ICDMW 2010)

**Sydney, Australia
13 December 2010**

Pages 1-694



**IEEE Catalog Number: CFP1056B-PRT
ISBN: 978-1-4244-9244-2**

2010 IEEE International Conference on Data Mining Workshops

ICDM-W 2010

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

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

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 |
| <i>Hardy Kremer, Stephan Günemann, and Thomas Seidl</i> | |
| A New Data Mining Model for Hurricane Intensity Prediction | 98 |
| <i>Yu Su, Sudheer Chelluboina, Michael Hahsler, and Margaret H. Dunham</i> | |
| Mining Spatial Co-occurrence of Drought Events from Climate Data of India | 106 |
| <i>Kollukuduru Sravanthi and K.S. Rajan</i> | |
| Data Analysis in Los Angeles Long Beach with Seasonal Time Series Model | 113 |
| <i>Weiqiang Wang and Zhendong Niu</i> | |

Knowledge Discovery Using Cloud and Distributed Computing Platforms—KDCloud 2010

| | |
|--|-----|
| On Finding Similar Items in a Stream of Transactions | 121 |
| <i>Andrea Campagna and Rasmus Pagh</i> | |
| Minimum Spanning Tree Based Classification Model for Massive Data with MapReduce Implementation | 129 |
| <i>Jin Chang, Jun Luo, Joshua Zhexue Huang, Shengzhong Feng, and Jianping Fan</i> | |
| Distributed, Scalable Clustering for Detecting Halos in Terascale Astronomy Datasets | 138 |
| <i>Srivatsava Daruru, Sankari Dhandapani, Gunjan Gupta, Ilian Iliev, Weijia Xu, Paul Navratil, Nena Marín, and Joydeep Ghosh</i> | |
| Challenges in Scheduling Aggregation in Cyberphysical Information Processing Systems | 148 |
| <i>James L. Horey</i> | |
| Parallel User Profiling Based on Folksonomy for Large Scaled Recommender Systems: An Implimentation of Cascading MapReduce | 154 |
| <i>Huizhi Liang, Jim Hogan, and Yue Xu</i> | |
| dSimpleGraph: A Novel Distributed Clustering Algorithm for Exploring Very Large Scale Unknown Data Sets | 162 |
| <i>Li Lu, Yunhong Gu, and Robert Grossman</i> | |
| S4: Distributed Stream Computing Platform | 170 |
| <i>Leonardo Neumeyer, Bruce Robbins, Anish Nair, and Anand Kesari</i> | |
| Parallel EM-Clustering: Fast Convergence by Asynchronous Model Updates | 178 |
| <i>Claudia Plant and Christian Böhm</i> | |
| Parallelizing an Information Theoretic Co-clustering Algorithm Using a Cloud Middleware | 186 |
| <i>Venkatram Ramanathan, Wenjing Ma, Vignesh T. Ravi, Tantan Liu, and Gagan Agrawal</i> | |

Spatial and Spatio-temporal Data Mining—SSTDM 2010

| | |
|--|-----|
| Putting Pixels in Place: A Storage Layout Language for Scientific Data | 194 |
| <i>Peter Baumann, Shams Feyzabadi, and Constantin Jucovschi</i> | |
| Using Time Series Segmentation for Deriving Vegetation Phenology Indices from MODIS NDVI Data | 202 |
| <i>Varun Chandola, Dafeng Hui, Lianhong Gu, Budhendra Bhaduri, and Ranga Raju Vatsavai</i> | |
| Comparing Vessel Trajectories Using Geographical Domain Knowledge and Alignments | 209 |
| <i>Gerben K.D. de Vries, Willem Robert van Hage, and Maarten van Someren</i> | |
| Learning Maximum Lag for Grouped Graphical Granger Models | 217 |
| <i>Amit Dhurandhar</i> | |
| 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 |
| <i>Sebnem Düzgün and Seyma Özcan Yavuzoglu</i> | |
| Computing Popular Places Using Graphics Processors | 233 |
| <i>Marta Fort, J. Antoni Sellarès, and Nacho Valladares</i> | |
| Sample Bias due to Missing Data in Mobility Surveys | 241 |
| <i>Dirk Hecker, Hendrik Stange, Christine Körner, and Michael May</i> | |
| Spatio-Temporal Symbolization of Multidimensional Time Series | 249 |
| <i>Shohei Hidaka and Chen Yu</i> | |
| Curvature Maxima-based Trajectories Mining | 257 |
| <i>Shoji Hirano and Shusaku Tsumoto</i> | |
| A K-Main Routes Approach to Spatial Network Activity Summarization: A Summary of Results | 265 |
| <i>Dev Oliver, Abdussalam Bannur, James M. Kang, Shashi Shekhar, and Renee Bousseilaire</i> | |
| Unsupervised Semantic Labeling Framework for Identification of Complex Facilities in High-Resolution Remote Sensing Images | 273 |
| <i>Ranga Raju Vatsavai, Anil Cheriyyadat, and Shaun Gleason</i> | |
| Measuring Similarity for Multidimensional Sequences | 281 |
| <i>Hui Wang, Zhiwei Lin, Sally McClean, and Jun Liu</i> | |

Social Interactions Analysis and Services Providers—SIASP 2010

| | |
|--|-----|
| Towards Community Discovery in Signed Collaborative Interaction Networks | 288 |
| <i>Petko Bogdanov, Nicholas D. Larusso, and Ambuj Singh</i> | |
| Catching a Viral Video | 296 |
| <i>Tom Broxton, Yannet Interian, Jon Vaver, and Mirjam Wattenhofer</i> | |

| | |
|---|-----|
| Improving Matching Process in Social Network | 305 |
| <i>Lin Chen, Richi Nayak, and Yue Xu</i> | |
| Extracting Representative Tags for Flickr Users | 312 |
| <i>Xian Chen and Hyoseop Shin</i> | |
| Detecting and Tracking Community Dynamics in Evolutionary Networks | 318 |
| <i>Zhengzhang Chen, Kevin A. Wilson, Ye Jin, William Hendrix, and Nagiza F. Samatova</i> | |
| The Meme Ranking Problem: Maximizing Microblogging Virality | 328 |
| <i>Dino Ienco, Francesco Bonchi, and Carlos Castillo</i> | |
| Modeling and Comparing the Influence of Neighbors on the Behavior of Users in Social and Similarity Networks | 336 |
| <i>Mohsen Jamali and Martin Ester</i> | |
| Modeling the Temporal Dynamics of Social Rating Networks Using Bidirectional Effects of Social Relations and Rating Patterns | 344 |
| <i>Mohsen Jamali, Gholamreza Haffari, and Martin Ester</i> | |
| A Social Matching System for an Online Dating Network: A Preliminary Study | 352 |
| <i>Richi Nayak, Meng Zhang, and Lin Chen</i> | |
| RnR: Extracting Rationale from Online Reviews and Ratings | 358 |
| <i>Dwi AP Rahayu, Shonali Krishnaswamy, Oshadi Alahakoon, and Cyril Labbe</i> | |
| Bridging Folksonomies and Domain Ontologies: Getting Out Non-taxonomic Relations | 369 |
| <i>Chiraz Trabelsi, Aicha Ben Jrad, and Sadok Ben Yahia</i> | |
| Toward Finding Hidden Communities Based on User Profile | 380 |
| <i>Tetsuya Yoshida</i> | |
| Visual Analytics and Knowledge Discovery—VAKD 2010 | |
| A Visual Analytics Tool for Analysing Microarray Data | 388 |
| <i>Martin Amélie, Monique Noirhomme-Fraiture, Quang Vinh Nguyen, and Simeon Simoff</i> | |
| Analysing the Large-Scale, Time-Oriented, VAST 2010 MC2 Dataset with PRISMA | 396 |
| <i>Rafael Guimarães, Nikolas Carneiro, Anderson Marques, Bianchi Meiguins, and Aruanda Simões Meiguins</i> | |
| Evolving Ensemble-Clustering to a Feedback-Driven Process | 401 |
| <i>Martin Hahmann, Dirk Habich, and Wolfgang Lehner</i> | |
| Visually Controllable Data Mining Methods | 409 |
| <i>Kai Puolamäki, Panagiotis Papapetrou, and Jefrey Lijffijt</i> | |
| User-Based Active Learning | 418 |
| <i>Christin Seifert and Michael Granitzer</i> | |
| SIMPLE: Interactive Analytics on Patent Data | 426 |
| <i>Scott Spangler, Ying Chen, Jeffrey Kreulen, Stephen Boyer, Thomas Griffin, Alfredo Alba, Linda Kato, Ana Lelescu, and Su Yan</i> | |

| | |
|---|-----|
| High-Dimensional Multimodal Distribution Embedding | 434 |
| <i>Eniko Szekely, Eric Bruno, and Stephane Marchand-Maillet</i> | |

| | |
|---|-----|
| BODY—Buckets of Disease Symptoms for Disease Outbreak Analysis | 442 |
| <i>Hanisha Veeramachaneni, Soujanya Vadapalli, and Kamalakar Karlapalem</i> | |

Privacy Aspects of Data Mining—PADM 2010

| | |
|--|-----|
| Language-Based Enforcement of Access Constraints on Geoscientific Image Data | 450 |
| <i>Peter Baumann</i> | |

| | |
|--|-----|
| Identifying and Preventing Data Leakage in Multi-relational Classification | 458 |
| <i>Hongyu Guo, Herna L. Viktor, and Eric Paquet</i> | |

| | |
|---|-----|
| On Attribute Disclosure in Randomization Based Privacy Preserving Data Publishing | 466 |
| <i>Ling Guo, Xiaowei Ying, and Xintao Wu</i> | |

| | |
|--|-----|
| Privacy Violations Using Microtargeted Ads: A Case Study | 474 |
| <i>Aleksandra Korolova</i> | |

| | |
|--|-----|
| Frequent Closed Itemset Mining with Privacy Preserving for Distributed Databases | 483 |
| <i>Shin-ya Kuno, Koichiro Doi, and Akihiro Yamamoto</i> | |

| | |
|---|-----|
| Anonymizing Graphs Against Weight-Based Attacks | 491 |
| <i>Yidong Li and Hong Shen</i> | |

| | |
|--|-----|
| A Privacy Preserving Framework for Gaussian Mixture Models | 499 |
| <i>Madhusudana Shashanka</i> | |

| | |
|--|-----|
| Identifying Similar Neighborhood Structures in Private Social Networks | 507 |
| <i>Lisa Singh and Clare Schramm</i> | |

Semantic Aspects in Data Mining—SADM 2010

| | |
|--|-----|
| Geospatial Schema Matching with High-Quality Cluster Assurance and Location Mining from Social Network | 517 |
| <i>Latifur Khan, Jeffrey Partyka, Satyen Abrol, and Bhavani Thuraisingham</i> | |

| | |
|---|-----|
| Semantic Content Filtering with Wikipedia and Ontologies | 518 |
| <i>Pekka Malo, Pyry Siitari, Oskar Ahlgren, Jyrki Wallenius, and Pekka Korhonen</i> | |

| | |
|---|-----|
| From Pattern Discovery to Pattern Interpretation in Movement Data | 527 |
| <i>Rebecca Ong, Monica Wachowicz, Mirco Nanni, and Chiara Renso</i> | |

| | |
|---|-----|
| Incorporating Multi-partite Networks and Expertise to Construct Related-Term Graphs | 535 |
| <i>Jyh-Ren Shieh, Ching-Yung Lin, Shun-Xuan Wang, Yung-Huan Hsieh, and Ja-Ling Wu</i> | |

| | |
|--|-----|
| Analysis of Collaborative Writing Processes Using Hidden Markov Models and Semantic Heuristics | 543 |
| <i>Vilaythong Southavilay, Kalina Yacef, and Rafael A. Calvo</i> | |

| | |
|---|-----|
| Using Ontology and Sequence Information for Extracting Behavior Patterns from Web Navigation Logs | 549 |
| <i>Hakan Yilmaz and Pinar Senkul</i> | |

Applications of Data Mining and Modeling in Government and Industry—ADMMGI 2010

| | |
|--|-----|
| Variable Productivity Adjustment Estimation for Function Point Project Delivery | 557 |
| <i>Saeed Bagheri, Krishna Ratakonda, and Rakesh Mohan</i> | |
| Graphics Classification for Enterprise Knowledge Management | 562 |
| <i>Divna Djordjevic and Rayid Ghani</i> | |
| Efficient Rule Generation for Dominant Class Problems on LARM | 570 |
| <i>JuiHsi Fu and SingLing Lee</i> | |
| Application of Data Mining for Anti-money Laundering Detection: A Case Study | 577 |
| <i>Nhien An Le Khac and M-Tahar Kechadi</i> | |
| Empirical Analysis: News Impact on Stock Prices Based on News Density | 585 |
| <i>Xiaodong Li, Xiaotie Deng, Feng Wang, and Keren Dong</i> | |
| Sales Forecasting of IT Products Using a Hybrid MARS and SVR Model | 593 |
| <i>Chi-Jie Lu, Tian-Shyug Lee, and Chia-Mei Lian</i> | |
| Unsupervised DRG Upcoding Detection in Healthcare Databases | 600 |
| <i>Wei Luo and Marcus Gallagher</i> | |
| Mining Heterogeneous ADS-B Data Sets for Probabilistic Models of Pilot Behavior | 606 |
| <i>Ron Marsh and Kirk Ogaard</i> | |
| Detecting Non-compliant Consumers in Spatio-Temporal Health Data: A Case Study from Medicare Australia | 613 |
| <i>K.S. Ng, Y. Shan, D.W. Murray, A. Sutinen, B. Schwarz, D. Jeacocke, and J. Farrugia</i> | |
| Visual Predictions of Currency Crises Using Self-Organizing Maps | 623 |
| <i>Peter Sarlin and Dorina Marghescu</i> | |
| Gaussian Processes for Dispatching Rule Selection in Production Scheduling: Comparison of Learning Techniques | 631 |
| <i>Bernd Scholz-Reiter, Jens Heger, and Torsten Hildebrandt</i> | |
| Using SOM-Ward Clustering and Predictive Analytics for Conducting Customer Segmentation | 639 |
| <i>Zhiyuan Yao, Tomas Eklund, and Barbro Back</i> | |
| Quantum Path Integral Inspired Query Sequence Suggestion for User Search Task Simplification | 647 |
| <i>Baojun Yue, Jun Yan, Heng Liang, Ning Liu, Lei Ji, Fengshan Bai, and Zheng Chen</i> | |
| Less Effort, More Outcomes: Optimising Debt Recovery with Decision Trees | 655 |
| <i>Yanchang Zhao, Hans Bohlscheid, Shanshan Wu, and Longbing Cao</i> | |

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 |
| <i>Farit Mochamad Afendi, Latifah K. Darusman, Aki Hirai, Md. Altaf-Ul-Amin, Hiroki Takahashi, Kensuke Nakamura, and Shigehiko Kanaya</i> | |
| The Multiple Alignment Algorithm for Metabolic Pathways without Abstraction | 669 |
| <i>Wenbin Chen, Andrea M. Rocha, William Hendrix, Matthew Schmidt, and Nagiza F. Samatova</i> | |
| Fixed-Parameter Tractable Combinatorial Algorithms for Metabolic Networks Alignments | 679 |
| <i>Qiong Cheng, Jinpeng Wei, Alexander Zelikovsky, and Mitsunori Ogihara</i> | |
| Efficient Alignments of Metabolic Networks with Bounded Treewidth | 687 |
| <i>Qiong Cheng, Piotr Berman, Rob Harrison, and Alex Zelikovsky</i> | |
| Curve Profiling Feature: Novel Compact Representation for <i>Drosophila</i> Embryonic Gene Expression Pattern Mining | 695 |
| <i>Chunsheng Fang, Minlu Zhang, Anca L. Ralescu, and Jason L. Lu</i> | |
| A Sequence Data Mining Protocol to Identify Best Representative Sequence for Protein Domain Families | 703 |
| <i>V.S. Gowri, Khader Shameer, Chilamakuri Chandra Sekhar Reddy, Prashant Shingate, and Ramanathan Sowdhamini</i> | |
| Parametric Templates: A New Enzyme Active-Site Prediction Algorithm | 711 |
| <i>Tsuyoshi Kato, Kazuhiro Suwa, and Nozomi Nagano</i> | |
| Overcoming the Curse of Dimensionality in a Statistical Geometry Based Computational Protein Mutagenesis | 719 |
| <i>Majid Masso</i> | |
| Interaction as an Interestingness Measure | 726 |
| <i>Martin McGrane and Simon K. Poon</i> | |
| Is Topotecan Effective at Killing Cancer Cells? | 732 |
| <i>Ricardo Santiago-Mozos, Imtiaz A. Khan, and Michael G. Madden</i> | |
| Cost-Sensitive Feature Selection Based on the Set Covering Machine | 740 |
| <i>Raúl Santos-Rodríguez and Darío García-García</i> | |
| Combining Time Series Similarity with Density-Based Clustering to Identify Fiber Bundles in the Human Brain | 747 |
| <i>Junming Shao, Klaus Hahn, Qinli Yang, Christian Böhm, Afra Wohlschläger, Nicholas Myers, and Claudia Plant</i> | |
| Evaluation of Protein Backbone Alphabets: Using Predicted Local Structure for Fold Recognition | 755 |
| <i>Kyong Jin Shim</i> | |

| | |
|--|-----|
| Fast Anomaly Detection in Dynamic Clinical Datasets Using Near-Optimal Hashing with Concentric Expansions | 763 |
| <i>Zeeshan Syed and Ilan Rubinfeld</i> | |
| An Empirical Comparison of Platt Calibration and Inductive Confidence Machines for Predictions in Drug Discovery | 771 |
| <i>Nikil Dale</i> | |
| FDCluster: Mining Frequent Closed Discriminative Bicluster without Candidate Maintenance in Multiple Microarray Datasets | 779 |
| <i>Miao Wang, Xuequn Shang, Shaohua Zhang, and Zhanhuai Li</i> | |
| An Integrative Scoring Approach to Identify Transcriptional Regulations Controlling Lung Surfactant Homeostasis | 787 |
| <i>Minlu Zhang, Chunsheng Fang, Yan Xu, Raj K. Bhatnagar, and Long J. Lu</i> | |
| International Workshop on Chance Discovery—IWCD 2010 | |
| Curation in Chance Discovery | 793 |
| <i>Akinori Abe</i> | |
| Text Categorization with Considering Temporal Patterns of Term Usages | 800 |
| <i>Hidenao Abe and Shusaku Tsumoto</i> | |
| Paired Evaluators Method to Track Concept Drift: An Application for Hedge Funds Operations | 808 |
| <i>Masabumi Furuhashi, Takanobu Mizuta, and Jihei So</i> | |
| Innovators Marketplace as Game on the Table versus Board on the Web | 816 |
| <i>Yukio Ohsawa, Kensuke Okamoto, Yuma Takahashi, and Yoko Nishihara</i> | |
| Finding Rare Patterns with Weak Correlation Constraint | 822 |
| <i>Yoshiaki Okubo, Makoto Haraguchi, and Takeshi Nakajima</i> | |
| Temporal Logic $TL \wedge Z_{\{DU\}}$ Modeling Local and Global Discovery with Logical Uncertainty | 830 |
| <i>Vladimir V. Rybakov</i> | |
| Automated Empirical Selection of Rule Induction Methods Based on Recursive Iteration of Resampling Methods and Multiple Testing | 835 |
| <i>Shusaku Tsumoto and Shoji Hirano</i> | |
| Change with Delayed Labeling: When is it Detectable? | 843 |
| <i>Indrė Žliobaitė</i> | |
| Reliability Issues in Knowledge Discovery—RIKD 2010 | |
| Towards a Reliable Framework of Uncertainty-Based Group Decision Support System | 851 |
| <i>Junyi Chai and James N.K. Liu</i> | |

| | |
|--|-----|
| A Study on the Accuracy of Frequency Measures and Its Impact on Knowledge Discovery in Single Sequences | 859 |
| <i>Min Gan and Honghua Dai</i> | |
| Co-occurrence Based Predictors for Estimating Query Difficulty | 867 |
| <i>Hazra Imran and Aditi Sharan</i> | |
| Batch Mode Sparse Active Learning | 875 |
| <i>Lixin Shi and Yuhang Zhao</i> | |
| Towards the Detection of Potential Contradictions in Fuzzy Ontology Using a High Level Net Approach Integrated with Uncertainty Inference | 883 |
| <i>Ke Wang, James N.K. Liu, and Wei-Min Ma</i> | |
| Learning Robust Bayesian Network Classifiers in the Space of Markov Equivalent Classes | 891 |
| <i>Zhongfeng Wang, Zhihai Wang, and Bin Fu</i> | |
| Learning Restricted Bayesian Network Classifiers with Mixed Non-i.i.d. Sampling | 899 |
| <i>Zhongfeng Wang, Zhihai Wang, and Bin Fu</i> | |
| Semantic Image Retrieval Based on Multiple-Instance Learning | 905 |
| <i>Min Yao, Wenshen Yi, Rong Zhu, and Ran Cheng</i> | |
| Domain Driven Data Mining—DDDM 2010 | |
| Link Prediction Across Multiple Social Networks | 911 |
| <i>Muhammad Aurangzeb Ahmad, Zoheb Borbora, Jaideep Srivastava, and Noshir Contractor</i> | |
| Controlling Consistency in Top-N Recommender Systems | 919 |
| <i>Paolo Cremonesi and Roberto Turrin</i> | |
| Evaluating Association Rules by Quantitative Pairwise Property Comparisons | 927 |
| <i>Elnaz Delpisheh and John Z. Zhang</i> | |
| Trading Tests of Long-Term Market Forecast by Text Mining | 935 |
| <i>Kiyoshi Izumi, Takashi Goto, and Tohgoroh Matsui</i> | |
| Network Anomaly Detection Using a Commute Distance Based Approach | 943 |
| <i>Nguyen Lu Dang Khoa, Tahereh Babaie, Sanjay Chawla, and Zainab Zaidi</i> | |
| Vote-Based LELC for Positive and Unlabeled Textual Data Streams | 951 |
| <i>Bo Liu, Yanshan Xiao, Longbing Cao, and Philip S. Yu</i> | |
| Domain-Driven Data Mining for IT Infrastructure Support | 959 |
| <i>Girish Keshav Palshikar, Harrick M. Vin, Mohammed Mudassar, and Maitreya Natu</i> | |
| Insights from Applying Sequential Pattern Mining to E-commerce Click Stream Data | 967 |
| <i>Arthur Pitman and Markus Zanker</i> | |
| Medically Driven Data Mining Application: Recognition of Health Problems from Gait Patterns of Elderly | 976 |
| <i>Bogdan Pogorelc and Matjaž Gams</i> | |

| | |
|--|------|
| Contextual Sequential Pattern Mining | 981 |
| <i>Julien Rabatel, Sandra Bringay, and Pascal Poncelet</i> | |
| Using Taxonomies to Perform Aggregated Querying over Imprecise Data | 989 |
| <i>Atanu Roy, Chandrima Sarkar, and Rafal A. Angryk</i> | |
| Sequence Alignment Based Analysis of Player Behavior in Massively Multiplayer Online Role-Playing Games (MMORPGs) | 997 |
| <i>Kyong Jin Shim and Jaideep Srivastava</i> | |
| Agent Assignment for Process Management: Agent Performance Evaluation Framework | 1005 |
| <i>Ramzan Talib, Bernhard Volz, and Stefan Jablonski</i> | |
| The Effectiveness of a New Negative Correlation Learning Algorithm for Classification Ensembles | 1013 |
| <i>Shuo Wang and Xin Yao</i> | |
| Learning with Actionable Attributes: Attention—Boundary Cases! | 1021 |
| <i>Indre Žliobaite and Mykola Pechenizkiy</i> | |
| Data Mining for Service—DMS 2010 | |
| Large-Scale Customized Models for Advertisers | 1029 |
| <i>Abraham Bagherjeiran, Andrew Hatch, Adwait Ratnaparkhi, and Rajesh Parekh</i> | |
| Knowledge File System—A Principled Approach to Personal Information Management | 1037 |
| <i>Kuiyu Chang, I. Wayan Tresna Perdana, Bramandia Ramadhana, Kailash Sethuraman, Truc Viet Le, and Neha Chachra</i> | |
| Automated Prompting in a Smart Home Environment | 1045 |
| <i>Barnan Das, Chao Chen, Nairanjana Dasgupta, Diane J. Cook, and Adriyana M. Seelye</i> | |
| Contextual Segmentation: Using Context to Improve Behavior Predictive Models in E-commerce | 1053 |
| <i>Maria Francesca Faraone, Michele Gorgoglione, and Cosimo Palmisano</i> | |
| Interesting Subset Discovery and Its Application on Service Processes | 1061 |
| <i>Maitreya Natu and Girish Keshav Palshikar</i> | |
| Discovering Temporal Features and Relations of Activity Patterns | 1069 |
| <i>Ehsan Nazerfard, Parisa Rashidi, and Diane J. Cook</i> | |
| Towards Data-Oriented Hospital Services: Data Mining-Based Hospital Management | 1076 |
| <i>Shusaku Tsumoto, Shoji Hirano, and Yuko Tsumoto</i> | |

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 |
| <i>Noraswaliza Abdullah, Yue Xu, Shlomo Geva, and Jinghong Chen</i> | |
| The PARIS Algorithm for Determining Latent Topics | 1092 |
| <i>Michal Aharon, Ira Cohen, Arik Itskovitch, Inbal Marhaim, and Ron Banner</i> | |
| A Block Mixture Model for Pattern Discovery in Preference Data | 1100 |
| <i>Nicola Barbieri, Massimo Guarascio, and Giuseppe Manco</i> | |
| Mining Arabic Business Reviews | 1108 |
| <i>Mohamed Elhawary and Mohamed Elfeky</i> | |
| Sentence-Level and Document-Level Sentiment Mining for Arabic Texts | 1114 |
| <i>Noura Farra, Elie Challita, Rawad Abou Assi, and Hazem Hajj</i> | |
| High-Order Concept Associations Mining and Inferential Language Modeling for Online Review Spam Detection | 1120 |
| <i>C.L. Lai, K.Q. Xu, Raymond Y.K. Lau, Yuefeng Li, and Dawei Song</i> | |
| Mining Users' Opinions Based on Item Folksonomy and Taxonomy for Personalized Recommender Systems | 1128 |
| <i>Huizhi Liang, Yue Xu, and Yuefeng Li</i> | |
| A Framework for Emotion Mining from Text in Online Social Networks | 1136 |
| <i>Mohamed Yassine and Hazem Hajj</i> | |
| Augmenting Chinese Online Video Recommendations by Using Virtual Ratings Predicted by Review Sentiment Classification | 1143 |
| <i>Weishi Zhang, Guiguang Ding, Li Chen, and Chunping Li</i> | |

Optimization Based Techniques for Emerging Data Mining Problems—OEDM 2010

| | |
|---|------|
| ValuePick: Towards a Value-Oriented Dual-Goal Recommender System | 1151 |
| <i>Leman Akoglu and Christos Faloutsos</i> | |
| Efficient Dimensionality Reduction on Undersampled Problems through Incremental Discriminative Common Vectors | 1159 |
| <i>Francesc J. Ferri, Katerine Díaz-Chito, and Wladimiro Díaz-Villanueva</i> | |
| XML Documents Clustering Using Tensor Space Model—A Preliminary Study | 1167 |
| <i>Sangetha Kutty, Richi Nayak, and Yuefeng Li</i> | |
| Ant Colony Optimization Algorithm Based on Immunity Vaccine and Dynamic Pheromone Updating | 1174 |
| <i>Wanjun Liu, Juan Zhang, and Junli Liu</i> | |
| Robust Low-Rank Subspace Segmentation with Semidefinite Guarantees | 1179 |
| <i>Yuzhao Ni, Ju Sun, Xiaotong Yuan, Shuicheng Yan, and Loong-Fah Cheong</i> | |

| | |
|--|------|
| Find Intelligent Knowledge by Second-Order Mining: Three Cases from China | 1189 |
| <i>Guangli Nie, Lingling Zhang, Yuejin Zhang, Wei Deng, and Yong Shi</i> | |
| Semi-supervised PLSA for Document Clustering | 1196 |
| <i>Lingfeng Niu and Yong Shi</i> | |
| Cluster Cores and Modularity Maximization | 1204 |
| <i>Michael Ovelgönne and Andreas Geyer-Schulz</i> | |
| Reviewer Profiling Using Sparse Matrix Regression | 1214 |
| <i>Evangelos E. Papalexakis, Nicholas D. Sidiropoulos, and Minos N. Garofalakis</i> | |
| Distributed Flow Algorithms for Scalable Similarity Visualization | 1220 |
| <i>Novi Quadrianto, Dale Schuurmans, and Alex J. Smola</i> | |
| Efficient Additive Models via the Generalized Lasso | 1228 |
| <i>Dimitri Semenovitch, Nobuyuki Morioka, and Arcot Sowmya</i> | |
| A Comparison of Objective Functions in Network Community Detection | 1234 |
| <i>Chuan Shi, Yanan Cai, Philip S. Yu, Zhenyu Yan, and Bin Wu</i> | |
| Large-Scale Matrix Factorization Using MapReduce | 1242 |
| <i>Zhengguo Sun, Tao Li, and Naphtali Rish</i> | |
| Integrating Symmetric Nonnegative Matrix Factorization and Normalized Cut Spectral Clustering | 1249 |
| <i>Zhichen Xia and Chris Ding</i> | |
| Integer Programming for Multi-class Active Learning | 1257 |
| <i>Dragomir Yankov, Suju Rajan, and Adwait Ratnaparkhi</i> | |
| ALPOS: A Machine Learning Approach for Analyzing Microblogging Data | 1265 |
| <i>Dan Zhang, Yan Liu, Richard D. Lawrence, and Vijil Chenthamarakshan</i> | |
| A Novel Ranking Algorithm for Service Matching Based on Agent Association Graphs | 1273 |
| <i>Hao Lan Zhang, Clement H.C. Leung, Gitesh K. Raikundalia, and Jing He</i> | |
| From Convex to Nonconvex: A Loss Function Analysis for Binary Classification | 1281 |
| <i>Lei Zhao, Musa Mammadov, and John Yearwood</i> | |
| Subspace Distance-Based Sampling Method for SVM | 1289 |
| <i>Xiaofei Zhou and Yong Shi</i> | |
| Large-Scale Analytics for Complex Instrumented Systems—LACIS 2010 | |
| Sliding HyperLogLog: Estimating Cardinality in a Data Stream over a Sliding Window | 1297 |
| <i>Yusra Chabchoub and Georges Hébrail</i> | |
| Using Adaptive Downsampling to Compare Time Series with Warping | 1304 |
| <i>Chih-Chun Chia and Zeeshan Syed</i> | |

| | |
|---|------|
| Multi-step Time Series Prediction in Complex Instrumented Domains | 1312 |
| <i>Amit Dhurandhar</i> | |
| dMaximalCliques: A Distributed Algorithm for Enumerating All Maximal Cliques and Maximal Clique Distribution | 1320 |
| <i>Li Lu, Yunhong Gu, and Robert Grossman</i> | |
| EigenDiagnostics: Spotting Connection Patterns and Outliers in Large Graphs | 1328 |
| <i>Koji Maruhashi and Christos Faloutsos</i> | |
| Enhancing Ubiquitous Systems through System Call Mining | 1338 |
| <i>Katharina Morik, Felix Jungermann, Nico Piatkowski, and Michael Engel</i> | |
| Parallelized Boosting with Map-Reduce | 1346 |
| <i>Indranil Palit and Chandan K. Reddy</i> | |

Contest Papers

| | |
|---|------|
| A Convex Combination of Models for Predicting Road Traffic | 1354 |
| <i>Carlos J. Gil Bellosta</i> | |
| Predicting Travel Times with Context-Dependent Random Forests by Modeling Local and Aggregate Traffic Flow | 1357 |
| <i>Benjamin Hamner</i> | |
| Predicting Future Traffic Congestion from Automated Traffic Recorder Readings with an Ensemble of Random Forests | 1360 |
| <i>Benjamin Hamner</i> | |
| Ensemble-Based Method for Task 2: Predicting Traffic Jam | 1363 |
| <i>Jingrui He, Qing He, Grzegorz Swirszcz, Yiannis Kamarianakis, Rick Lawrence, Wei Shen, and Laura Wynter</i> | |
| IEEE ICDM 2010 Contest: Traffic Prediction—Jams | 1366 |
| <i>Lukasz Romaszko</i> | |
| Traffic Velocity Prediction Using GPS Data: IEEE ICDM Contest Task 3 Report | 1369 |
| <i>Wei Shen, Yiannis Kamarianakis, Laura Wynter, Jingrui He, Qing He, Rick Lawrence, and Grzegorz Swirszcz</i> | |
| IEEE ICDM 2010 Contest: TomTom Traffic Prediction for Intelligent GPS Navigation | 1372 |
| <i>Marcin Wojnarski, Pawel Gora, Marcin Szczuka, Hung Son Nguyen, Joanna Swietlicka, and Demetris Zeinalipour</i> | |

Demo Papers

| | |
|--|------|
| Meerkat: Community Mining with Dynamic Social Networks | 1377 |
| <i>Jiyang Chen, Justin Fagnan, Randy Goebel, Reihaneh Rabbany, Farzad Sangi, Mansoureh Takaffoli, Eric Verbeek, and Osmar Zaiane</i> | |

| | |
|--|------|
| Clutter-Adaptive Visualization for Mobile Data Mining | 1381 |
| <i>Brett Gillick, Hasnain AlTair, Shonali Krishnaswamy, Jonathan Liono, Nicholas Nicoloudis, Abhijat Sinha, Arkady Zaslavsky, and Mohamed Medhat Gaber</i> | |
| Data Prediction Competitions—Far More than Just a Bit of Fun | 1385 |
| <i>Anthony Goldbloom</i> | |
| MCExplorer: Interactive Exploration of Multiple (Subspace) Clustering Solutions | 1387 |
| <i>Stephan Günnemann, Hardy Kremer, Ines Färber, and Thomas Seidl</i> | |
| Targeting Television Audiences Using Demographic Similarity | 1391 |
| <i>Brendan Kitts, Liang Wei, Dyng Au, Stefanie Zlomek, Ryan Brooks, and Brian Burdick</i> | |
| Clustering Performance on Evolving Data Streams: Assessing Algorithms and Evaluation Measures within MOA | 1400 |
| <i>Philipp Kranen, Hardy Kremer, Timm Jansen, Thomas Seidl, Albert Bifet, Geoff Holmes, and Bernhard Pfahringer</i> | |
| i-Analyst: An Agent-Based Distributed Data Mining Platform | 1404 |
| <i>Chayapol Moemeng, Xinhua Zhu, Longbing Cao, and Chen Jiahang</i> | |

Demo 2009

| | |
|---|------|
| Enhancing Document Exploration with OLAP | 1407 |
| <i>Zhibo Chen, Carlos Garcia-Alvarado, and Carlos Ordonez</i> | |
| QueRIE: A Query Recommender System Supporting Interactive Database Exploration | 1411 |
| <i>Sarika Mittal, Jothi Swarubini Vindhiya Varman, Gloria Chatzopoulou, Magdalini Eirinaki, and Neoklis Polyzotis</i> | |
| uRule: A Rule-Based Classification System for Uncertain Data | 1415 |
| <i>Biao Qin, Yuni Xia, Rakesh Sathyesh, Sunil Prabhakar, and Yicheng Tu</i> | |
| Adaptive Multimedia Mining on Distributed Stream Processing Systems | 1419 |
| <i>Deepak S. Turaga, Hyunggon Park, Rong Yan, and Olivier Verscheure</i> | |

Author Index