

ICDM Workshops 2007

Seventh IEEE International Conference on Data Mining – Workshops

Table of Contents

Welcome Message from the Conference Chairs	xiii
Preface	xiv
Workshops and Organizations	xvi

Data Mining in Web 2.0 Environments

Ask the Crowd to Find out What's Important	3
<i>Sisay Fissaha Adafre and Maarten de Rijke</i>	
Aspect Summarization from Blogosphere for Social Study	9
<i>Chia-Hui Chang and Kun-Chang Tsai</i>	
FiVaTech: Page-Level Web Data Extraction from Template Pages.....	15
<i>Mohammed Kayed, Chia-Hui Chang, Khaled Shaalan, and Moheb Ramzy Girgis</i>	
SOPS: Stock Prediction Using Web Sentiment	21
<i>Vivek Sehgal and Charles Song</i>	
HSN-PAM: Finding Hierarchical Probabilistic Groups from Large-Scale Networks	27
<i>Haizheng Zhang, Wei Li, Xuerui Wang, C. Lee Giles, Henry C. Foley, and John Yen</i>	
Extracting Author Meta-Data from Web Using Visual Features.....	33
<i>Shuyi Zheng, Ding Zhou, Jia Li, and C. Lee Giles</i>	

Knowledge Discovery from Multimedia Data and Multimedia Applications

Automatic Generation of Traditional Style Painting by Using Density-Based Color Clustering	41
<i>Guangyan Huang, Zhiming Ding, and Jing He</i>	
Semi-Automatic Semantic Annotation of Images	45
<i>Suzanne Little, Ovidio Salvetti, and Petra Pernert</i>	

Bit Sequences and Biclustering of Text Documents.....	51
<i>Selim Mimaroglu and Dan A. Simovici</i>	
Video Data Mining: Discovering Topics by Burst Detection in Video Streams.....	57
<i>Kimiaki Shirahama and Kuniaki Uehara</i>	
Tensor Space Learning for Analyzing Activity Patterns from Video Sequences	63
<i>Liang Wang, Christopher Leckie, Xiaozhe Wang, Ramamohanarao Kotagiri, and Jim Bezdek</i>	
Adapting SVM Classifiers to Data with Shifted Distributions	69
<i>Jun Yang, Rong Yan and Alexander G. Hauptmann</i>	

Mining and Management of Biological Data

A Comparative Study of Methods for Transductive Transfer Learning.....	77
<i>Andrew Arnold, Ramesh Nallapati, and William W. Cohen</i>	
Assessing Reliability of Protein-Protein Interactions by Semantic Data Integration.....	83
<i>Young-Rae Cho, Woochang Hwang, and Aidong Zhang</i>	
Characterizing RNA Secondary-Structure Features and Their Effects on Splice-Site Prediction	89
<i>Rezarta Islamaj Dogan, Lise Getoor, and W. John Wilbur</i>	
Statistical Approaches to Identifying Androgen Response Elements.....	95
<i>Li Li, Steffen Heber, Qiang Zhang, and Melvin E. Andersen</i>	
Mapping Gene/Protein Names in Free Text to Biomedical Databases	101
<i>Hongfang Liu, Manabu Torii, Zhang-zhi Hu, and Cathy Wu</i>	
A Content Based Pattern Analysis System for a Biological Specimen Collection.....	107
<i>Joyita Mallik, Ashok Samal, and Scott L. Gardner</i>	
Discovering Gene Expression Data from the Tables of Full Text Publications	113
<i>Brigitte Mathiak, Andreas Kupfer, Carolina Rio Bartulos, Tatjana Scope, Johann Weiland, and Silke Eckstein</i>	
Modeling and Management of Signal Transduction Pathways with Live Sequence Charts.....	119
<i>Claudia Täubner, Brigitte Mathiak, and Silke Eckstein</i>	

Data Mining in Medicine

Developing an Integrated Time-Series Data Mining Environment for Medical Data Mining.....	127
<i>Hidenao Abe, Hideto Yokoi, Miho Ohsaki, and Takahira Yamaguchi</i>	
Time-Annotated Sequences for Medical Data Mining	133
<i>Michele Berlingerio, Francesco Bonchi, Fosca Giannotti, and Franco Turini</i>	
Automatically Finding Images for Clinical Decision Support.....	139
<i>Dina Demner-Fushman, Sameer Antani, and George R. Thoma</i>	

Content-Based Support Environment (C-BASE): Data Preparation and Similarity Measurement	145
<i>Ameen Eetemadi, Mohammad-Reza Siadat, Hamid Soltanian-Zadeh, Farshad Fotouhi, and Kost Elisevich</i>	
Identifying Exacerbating Cases in Chronic Diseases Based on the Cluster Analysis of Trajectory Data on Laboratory Examinations	151
<i>Shoji Hirano and Shusaku Tsumoto</i>	
Predictive Data Mining for Lung Nodule Interpretation	157
<i>William Horsthemke, Ekarin Varutbangkul, Daniela Raicu, and Jacob Furst</i>	
Predictive Data Mining to Learn Health Vitals of a Resident in a Smart Home	163
<i>Vikramaditya Jakkula</i>	
Utilization of Data-Mining Techniques for Evaluation of Patterns of Asthma Drugs Use by Ambulatory Patients in a Large Health Maintenance Organization	169
<i>Mark Last, Rafael Carel, and Dotan Barak</i>	
Analysis of Relationship between Blood Stream Infection and Clinical Background in Patients' Lactobacillus Therapy by Data Mining	175
<i>Kimiko Matsuoka, Shigeki Yokoyama, Kunitomo Watanabe, and Shusaku Tsumoto</i>	
Data Modeling for Content-Based Support Environment (C-BASE): Application on Epilepsy Data Mining	181
<i>Mohammad-Reza Siadat, Hamid Soltanian-Zadeh, Farshad Fotouhi, Ameen Eetemadi, and Kost Elisevich</i>	

Optimization-Based Data Mining Techniques with Applications

Data Clustering with a Relational Push-Pull Model	189
<i>Adam Anthony and Marie desJardins</i>	
Learning What Makes a Society Tick	195
<i>Hung-Ching (Justin) Chen, Mark Goldberg, Malik Magdon-Ismael, and William A. Wallace</i>	
Distance Metric Learning through Optimization of Ranking	201
<i>Kreshna Gopal and Thomas R. Ioerger</i>	
Using Data Mining to Estimate Missing Sensor Data	207
<i>Le Gruenwald, Hamed Chok, and Mazen Aboukhamis</i>	
Cluster Analysis and Optimization in Color-Based Clustering for Image Abstract	213
<i>Jing He, Guangyan Huang, Yanchun Zhang, and Yong Shi</i>	
Predicting and Optimizing Classifier Utility with the Power Law	219
<i>Mark Last</i>	
Mining Distance-Based Outliers from Categorical Data	225
<i>Shuxin Li, Robert Lee, and Sheau-Dong Lang</i>	
Feature Selection for Nonlinear Kernel Support Vector Machines	231
<i>Olvi L. Mangasarian and Edward W. Wild</i>	

Categorical Attribute Transformation Technique for Multiple Criteria Quadratic Programming Classification Model	237
<i>Yi Peng and Gang Kou</i>	
Physical Analysis of Precipitation Factors Based on SVM Method	243
<i>Xiong Qiufen, Gao Jie, Liu Huanzhu, and Shao Mingxuan</i>	
An Efficient Fitness Assignment Based on Dominating Tree	247
<i>Chuan Shi, Zhongzhi Shi, and Bin Wu</i>	
A Regularized Multiple Criteria Linear Program for Classification	253
<i>Yong Shi, Yingjie Tian, Xiaojun Chen, and Peng Zhang</i>	
Dual Fuzzy-Possibilistic Co-clustering for Document Categorization	259
<i>William-Chandra Tjhi and Lihui Chen</i>	
Generalized Additive Models from a Neural Network Perspective	265
<i>D. A. de Waal and J. V. du Toit</i>	
A Novel Rule Weighting Approach in Classification Association Rule Mining	271
<i>Yanbo J. Wang, Qin Xin, and Frans Coenen</i>	
Semi-supervised Kernel Logistic Regression and Its Extension to Active Learning Based on A-Optimality	277
<i>Yasutoshi Yajima, and Teppei Sato</i>	
Classification with Choquet Integral with Respect to Signed Non-additive Measure	283
<i>Nian Yan, Zhenyuan Wang, and Zhengxin Chen</i>	
Multiple-Criteria Linear Programming for VIP E-Mail Behavior Analysis	289
<i>Peng Zhang and Jinran Dai</i>	

High Performance Data Mining

A Novel Parallel Boolean Approach for Discovering Frequent Itemsets	297
<i>Wassim Ayadi and Khedija Arour</i>	
Collaborative Filtering Using Orthogonal Nonnegative Matrix Tri-factorization	303
<i>Gang Chen, Fei Wang, and Changshui Zhang</i>	
Extracting Knowledge to Predict TSP Asymptotic Time Complexity	309
<i>Paula Cecilia Fritzsche, Dolores Rexachs, and Emilio Luque</i>	
Twin Kernel Embedding with Back Constraints	319
<i>Yi Guo, Paul W. Kwan, and Junbin Gao</i>	
An Efficient Technique for Mining Approximately Frequent Substring Patterns	325
<i>Xiaonan Ji and James Bailey</i>	
Robust Unsupervised and Semisupervised Bounded C-Support Vector Machines	331
<i>Zhao Kun, Tian Ying-jie, and Deng Nai-yang</i>	
Large PSD Matrix Estimation from Partial Elements	337
<i>Hiroshi Kuwajima and Takashi Washio</i>	

Sparse Word Graphs: A Scalable Algorithm for Capturing Word Correlations in Topic Models	343
<i>Ramesh Nallapati, Amr Ahmed, William Cohen, and Eric Xing</i>	
Parallelized Variational EM for Latent Dirichlet Allocation: An Experimental Evaluation of Speed and Scalability.....	349
<i>Ramesh Nallapati, William Cohen, and John Lafferty</i>	
WC-Clustering: Hierarchical Clustering Using the Weighted Confidence Affinity Measure	355
<i>Baoying Wang and Imad Rahal</i>	
Semi-supervised Clustering Using Bayesian Regularization.....	361
<i>Zuobing Xu, Ram Akella, Mike Ching, and Renjie Tang</i>	
Experimental Comparison of Feature Subset Selection Methods	367
<i>Chulmin Yun and Jihoon Yang</i>	
Utility-Based Web Path Traversal Pattern Mining.....	373
<i>Lin Zhou, Ying Liu, Jing Wang, and Yong Shi</i>	

Mining Graphs and Complex Structures

Combining Collective Classification and Link Prediction	381
<i>Mustafa Bilgic, Galileo Mark Namata, and Lise Getoor</i>	
Simultaneous Heterogeneous Data Clustering Based on Higher Order Relationships.....	387
<i>Shouchun Chen, Fei Wang, and Changshui Zhang</i>	
Discovering Structural Anomalies in Graph-Based Data	393
<i>William Eberle and Lawrence Holder</i>	
Subgraph Support in a Single Large Graph.....	399
<i>Mathias Fiedler and Christian Borgelt</i>	
Tree Planar Languages.....	405
<i>Christophe Costa Florêncio</i>	
An Examination of Experimental Methodology for Classifiers of Relational Data	411
<i>Brian Gallagher and Tina Eliassi-Rad</i>	
GDClust: A Graph-Based Document Clustering Technique	417
<i>M. Shahriar Hossain and Rafal A. Angryk</i>	
Learning Term Dependency Links Using Information Theoretic Inclusion Measure	423
<i>Masoud Makrehchi and Mohamed S. Kamel</i>	
Exploiting Network Structure for Active Inference in Collective Classification.....	429
<i>Matthew J. Rattigan, Marc Maier, and David Jensen</i>	
Resume Mining of Communities in Social Network	435
<i>Bin Wu, Xin Pei, JianBin Tan, and Yi Wang</i>	
A Divisive Hierarchical Structural Clustering Algorithm for Networks.....	441
<i>Nurcan Yuruk, Mutlu Mete, Xiaowei Xu, and Thomas A. J. Schweiger</i>	

Data Mining on Uncertain Data

Counterpropagation Neural Network for Stochastic Conditional Simulation: An Application with Berea Sandstone.....	449
<i>Lance E. Besaw and Donna M. Rizzo</i>	
Genre Categorization of Web Pages	455
<i>Jebari Chaker and Ounelli Habib</i>	
Segmenting Multi-attribute Sequences Using Dynamic Bayesian Networks.....	465
<i>Robert Gwadera, Janne Toivola, and Jaakko Hollmén</i>	
Error-Aware Density-Based Clustering of Imprecise Measurement Values	471
<i>Dirk Habich, Peter B. Volk, Wolfgang Lehner, Ralf Dittmann, and Clemens Utzny</i>	
Skewed Class Distributions and Mislabeled Examples	477
<i>Jason Van Hulse, Taghi M. Khoshgoftaar, and Amri Napolitano</i>	
Reducing UK-Means to K-Means	483
<i>S. D. Lee, Ben Kao, and Reynold Cheng</i>	
Efficient Mining of Frequent Patterns from Uncertain Data	489
<i>Carson Kai-Sang Leung, Christopher L. Carmichael, and Boyu Hao</i>	
A Novel Ordering-Based Greedy Bayesian Network Learning Algorithm on Limited Data	495
<i>Feng Liu, Fengzhan Tian, and Qiliang Zhu</i>	
Granularity Conscious Modeling for Probabilistic Databases	501
<i>Eirinaios Michelakis, Daisy Zhe Wang, Minos Garofalakis, and Joseph M. Hellerstein</i>	
Representing Tuple and Attribute Uncertainty in Probabilistic Databases	507
<i>Prithviraj Sen, Amol Deshpande, and Lise Getoor</i>	
Targeting Input Data for Acoustic Bird Species Recognition Using Data Mining and HMMs	513
<i>Erika Vilches, Iván A. Escobar, Edgar E. Vallejo, and Charles E. Taylor</i>	
Incremental Integration of Probabilistic Models Learned from Data	519
<i>Jian Xu, Pedrito Maynard-Zhang, and Jianhua Chen</i>	

Data Streaming Mining and Management

Incremental Quantization for Aging Data Streams	527
<i>Fatih Altıparmak, David Chiu, and Hakan Ferhatosmanoglu</i>	
Hierarchical Classifier Combination and Its Application in Networks Intrusion Detection	533
<i>Morteza Analoui, Behrouz Minaei Bidgoli, and Mohammad Hossein Rezvani</i>	
An Approach for Incremental Semi-supervised SVM	539
<i>Wael Emara and Mehmed Kantardzic</i>	
Quality of Service-Driven Stream Mining	545
<i>Marcel Karnstedt, Kai-Uwe Sattler, Dirk Habich, and Wolfgang Lehner</i>	

Sequential Change Detection on Data Streams	551
<i>S. Muthukrishnan, Eric van den Berg, and Yihua Wu</i>	
Optimal Window Change Detection	557
<i>Jan Peter Patist</i>	
High-Speed Identification of Language and Script	563
<i>Alan Ratner and Ron Loui</i>	
Infrequent Item Mining in Multiple Data Streams	569
<i>Budhaditya Saha, Mihai Lazarescu, and Svetha Venkatesh</i>	
Stream Event Detection: A Unified Framework for Mining Outlier, Change and Burst Simultaneously over Data Stream	575
<i>Zhijian Yuan, Kai Du, Yan Jia, and Jiajia Miao</i>	
Toward Behavioral Modeling of a Grid System: Mining the Logging and Bookkeeping Files	581
<i>Xiangliang Zhang, Michèle Sebag, and Cécile Germain</i>	

Spatial and Spatio-temporal Data Mining

Space-Time Interpolation and Uncertainty Assessment of an Extreme Precipitation Index Using Geostatistical Cosimulation	589
<i>Ana Cristina Costa and Amílcar Soares</i>	
On Regional Association Rule Scoping	595
<i>Wei Ding, Christoph F. Eick, Xiaojing Yuan, Jing Wang, and Jean-Philippe Nicot</i>	
A Compact Representation of Spatio-temporal Data	601
<i>Sigal Elnekave, Mark Last, and Oded Maimon</i>	
Modeling Fundamental Geo-Raster Operations with Array Algebra	607
<i>Angelica Garcia Gutierrez and Peter Baumann</i>	
Pattern Mining as Abduction: From Snapshots to Spatio-temporal Sequential Patterns	613
<i>Shyamanta M. Hazarika</i>	
Query Expansion Using Topic and Location	619
<i>Shu Huang, Qiankun Zhao, Prasenjit Mitra, and C. Lee Giles</i>	
Knowledge Discovery in Entity Based Smart Environment Resident Data Using Temporal Relation Based Data Mining	625
<i>Vikramaditya R. Jakkula, Aaron S. Crandall, and Diane J. Cook</i>	
Space-Time Summarization of Multisensor Time Series. Case of Missing Data	631
<i>Marc Joliveau and Florian De Vuyst</i>	
Spatial Clustering Using the Likelihood Function	637
<i>April Kerby, David Marx, Ashok Samal, and Viacheslav Adamchuck</i>	
Diagnosing Similarity of Oscillation Trends in Time Series	643
<i>Leonardo E. Mariote, Claudia Bauzer Medeiros, and Ricardo da S. Torres</i>	

Areal Aggregated Crime Reasoning through Density Tracing.....	649
<i>Peter Phillips and Ickjai Lee</i>	
Formulating, Identifying and Analyzing Individual Spatial Knowledge.....	655
<i>Falko Schmid</i>	
Using Statistics and Spatial Data Mining to Study Land Cover in Wyoming: Can We Predict Vegetation Types from Environmental Variables?	661
<i>ZongBo Shang and Jeffery D. Hamerlinck</i>	
The Vegetation Outlook (VegOut): A New Tool for Providing Outlooks of General Vegetation Conditions Using Data Mining Techniques.....	667
<i>Tsegaye Tadesse and Brian Wardlow</i>	
A Hybrid Classification Scheme for Mining Multisource Geospatial Data	673
<i>Ranga Raju Vatsavai and Budhendra Bhaduri</i>	
Fast Mining of Complex Spatial Co-location Patterns Using GLIMIT	679
<i>Florian Verhein and Ghazi Al-Naymat</i>	
Spatio-temporal Analysis of the Relationship between South American Precipitation Extremes and the El Niño Southern Oscillation	685
<i>Elizabeth Wu and Sanjay Chawla</i>	

Privacy Aspects of Data Mining

Hiding Sensitive Trajectory Patterns.....	693
<i>Osman Abul, Maurizio Atzori, Francesco Bonchi, and Fosca Giannotti</i>	
Privacy-Preserving k-NN for Small and Large Data Sets	699
<i>Artak Amirbekyan and Vladimir Estivill-Castro</i>	
A Secure Clustering Algorithm for Distributed Data Streams	705
<i>Geetha Jagannathan, Krishnan Pillaipakkamnatt, and D. Umamo</i>	
Private Inference Control for Aggregate Database Queries	711
<i>Geetha Jagannathan and Rebecca N. Wright</i>	
Privacy-Preserving Data Mining Applications in the Malicious Model	717
<i>Murat Kantarcioglu and Onur Kardes</i>	
“Secure” Logistic Regression of Horizontally and Vertically Partitioned Distributed Databases	723
<i>Aleksandra B. Slavkovic, Yuval Nardi, and Matthew M. Tibbits</i>	
Simultaneous Pattern and Data Hiding in Unsupervised Learning.....	729
<i>Jie Wang, Jun Zhang, Lian Liu, and Dianwei Han</i>	
Author Index.....	735