Welcome Message from the Conference

Chairs

Welcome to ICDM 2008

Conference Organization

Steering Committee

Program Committee

Non-PC Reviewers

Corporate Sponsors

Invited Talks

Tutorials

Workshops

Panel Session

Regular Papers

On-line LDA: Adaptive Topic Models for Mining Text Streams with Applications to Topic Detection and Tracking ................................................................. 3
  Loulwah AlSumait, Daniel Barbará, and Carlotta Domeniconi
Unsupervised Cross-Domain Learning by Interaction Information Co-clustering .................................................. 13
  Shin Ando and Einoshin Suzuki
Paired Learners for Concept Drift ................................................................. 23
  Stephen H. Bach and Marcus A. Maloof
Predicting Future Decision Trees from Evolving Data ............................................. 33
  Mirko Böttcher, Martin Spott, and Rudolf Kruse
A Randomized Approach for Approximating the Number of Frequent Sets .................................................. 43
  Mario Boley and Henrik Grosskreutz
A Non-parametric Semi-supervised Discretization Method .................................................. 53
  A. Bondu, M. Boulle, V. Lemaire, S. Loiseau, and B. Duval
A Novel Language-Model-Based Approach for Image Object Mining and Re-ranking ........................................243
Jen-Hao Hsiao, Chu-Song Chen, and Ming-Syan Chen

Maximum Margin Clustering with Pairwise Constraints ..............................................................253
Yang Hu, Jingdong Wang, Nenghai Yu, and Xian-Sheng Hua

Collaborative Filtering for Implicit Feedback Datasets ........................................................................263
Yifan Hu, Yehuda Koren, and Chris Volinsky

Semi-supervised Learning from General Unlabeled Data .................................................................273
Kaizhu Huang, Zenglin Xu, Irwin King, and Michael R. Lyu

Metropolis Algorithms for Representative Subgraph Sampling .....................................................283
Christian Hübler, Hans-Peter Kriegel, Karsten Borgwardt, and Zoubin Ghahramani

Learning on Weighted Hypergraphs to Integrate Protein Interactions and Gene
Expressions for Cancer Outcome Prediction ......................................................................................293
TaeHyun Hwang, Ze Tian, Rui Kuangy, and Jean-Pierre Kocher

A Fast Method to Mine Frequent Subsequences from Graph Sequence Data ......................................303
Akihiro Inokuchi and Takashi Washio

Overlapping Matrix Pattern Visualization: A Hypergraph Approach ................................................313
Ruoming Jin, Yang Xiang, David Fuhry, and Feodor F. Dragan

A Robust Discriminative Term Weighting Based Linear Discriminant Method for Text
Classification .................................................................................................................................323
Khurum Nazir Junejo and Asim Karim

Clustering Uncertain Data Using Voronoi Diagrams ........................................................................333
Ben Kao, Sau Dan Lee, David W. Cheung, Wai-Shing Ho, and K. F. Chan

SCS: A New Similarity Measure for Categorical Sequences ............................................................343
Abdellali Kelil and Shengrui Wang

Toward Faster Nonnegative Matrix Factorization: A New Algorithm and Comparisons ..................353
Jingu Kim and Haesun Park

Scalable Tensor Decompositions for Multi-aspect Data Mining ......................................................363
Tamara G. Kolda and Jimeng Sun

Mining Periodic Behavior in Dynamic Social Networks ................................................................373
Mayank Lahiri and Tanya Y. Berger-Wolf

Unsupervised Face Annotation by Mining the Web .........................................................................383
Duy-Dinh Le and Shin’ichi Satoh

Border Sampling through Coupling Markov Chain Monte Carlo ...................................................393
Guichong Li, Nathalie Japkowicz, Trevor J. Stocki, and R. Kurt Ungar

Computationally Efficient Estimators for Dimension Reductions Using Stable Random
Projections ........................................................................................................................................403
Ping Li

Isolation Forest ...................................................................................................................................413
Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou

TEFE: A Time-Efficient Approach to Feature Extraction ................................................................423
Li-Ping Liu, Yang Yu, Yuan Jiang, and Zhi-Hua Zhou

Transductive Component Analysis ..................................................................................................433
Wei Liu, Dacheng Tao, and Jianzhuang Liu
Modeling and Predicting the Helpfulness of Online Reviews ................................................................. 443
   Yang Liu, Xiangji Huang, Aijun An, and Xiaohui Yu
LBF: A Labeled-Based Forecasting Algorithm and Its Application to Electricity Price Time Series ................................................................................................................................. 453
   Francisco Martínez–Álvarez, Alicia Troncoso, José C. Riquelme, and Jesús S. Aguilar–Ruiz
Enhancing the Stability of Spectral Ordering with Sparsification and Partial Supervision: Application to Paleontological Data .................................................................................................................. 462
   Dimitrios Mavroeidis and Ella Bingham
Scaling up Classifiers to Cloud Computers ........................................................................................................ 472
   Christopher Moretti, Karsten Steinhaeuser, Douglas Thain, and Nitesh V. Chawla
What Sperner Family Concept Class is Easy to Be Enumerated? ................................................................ 482
   Atsuyoshi Nakamura and Mineichi Kudo
Learning by Propagability .......................................................................................................................... 492
   Bingbing Ni, Shuicheng Yan, Ashraf Kassim, and Loong Fah Cheong
One-Class Collaborative Filtering .................................................................................................................. 502
   Rong Pan, Yunhong Zhou, Bin Cao, Nathan N. Liu, Rajan Lukose, Martin Scholz, and Qiang Yang
DisCo: Distributed Co-clustering with Map-Reduce: A Case Study towards Petabyte-Scale End-to-End Mining .......................................................................................................................... 512
   Spiros Papadimitriou and Jimeng Sun
Learning Bayesian Networks: A MAP Criterion for Joint Selection of Model Structure and Parameter .................................................................................................................................................. 522
   Carsten Riggelsen
Bayesian Co-clustering ................................................................................................................................. 530
   Hanhuai Shan and Arindam Banerjee
Temporal-Relational Classifiers for Prediction in Evolving Domains .................................................................. 540
   Umang Sharan and Jennifer Neville
xCrawl: A High-Recall Crawling Method for Web Mining ................................................................................. 550
   Kostyantyn Shchekotykhin, Dietmar Jannach, and Gerhard Friedrich
Comparison of Cluster Representations from Partial Second- to Full Fourth-Order Cross Moments for Data Stream Clustering ............................................................................................................ 560
   Mingzhou (Joe) Song and Lin Zhang
Web Mining for Understanding Stories through Graph Visualisation .................................................................. 570
   Ilija Subašić and Bettina Berendt
Balancing Spectral Clustering for Segmenting Spatio-temporal Observations of Multi-agent Systems ...................................................................................................................................................... 580
   Bálint Takács and Yiannis Demiris
Finding Good Itemsets by Packing Data ........................................................................................................ 588
   Nikolaj Tatti and Jilles Vreeken
Measuring Proximity on Graphs with Side Information ..................................................................................... 598
   Hanghang Tong, Huiming Qu, and Hani Jamjoom
Fast Counting of Triangles in Large Real Networks without Counting: Algorithms and Laws ................................................................. 608
  Charalampos E. Tsourakakis
Improving Collaborative Filtering Recommendations Using External Data ................................................................. 618
  Akhmed Umyarov and Alexander Tuzhilin
A Generative Probabilistic Model for Multi-label Classification ...................................................................................... 628
  Hongning Wang, Minlie Huang, and Xiaoyan Zhu
SpecVAT: Enhanced Visual Cluster Analysis .............................................................................................................. 638
  Liang Wang, Xin Geng, James Bezdek, Christopher Leckie, and Ramamohanarao Kotagiri
Dirichlet Process Based Evolutionary Clustering ......................................................................................................... 648
  Tianbing Xu, Zhongfei (Mark) Zhang, Philip S. Yu, and Bo Long
Evolutionary Clustering by Hierarchical Dirichlet Process with Hidden Markov State ............................................... 658
  Tianbing Xu, Zhongfei (Mark) Zhang, Philip S. Yu, and Bo Long
TOFA: Trace Oriented Feature Analysis in Text Categorization .................................................................................. 668
  Jun Yan, Ning Liu, Qiang Yang, Weiguo Fan, and Zheng Chen
Clustering Distributed Time Series in Sensor Networks .................................................................................................. 678
  Jie Yin and Mohamed Medhat Gaber
M3MIML: A Maximum Margin Method for Multi-instance Multi-label Learning ..................................................... 688
  Min-Ling Zhang and Zhi-Hua Zhou

Short Papers

RTM: Laws and a Recursive Generator for Weighted Time-Evolving Graphs ................................................................. 701
  Leman Akoglu, Mary McGlohon, and Christos Faloutsos
A Shrinkage Approach for Modeling Non-stationary Relational Autocorrelation .................................................... 707
  Pelin Angin and Jennifer Neville
Latent Dirichlet Allocation and Singular Value Decomposition Based Multi-document Summarization ......................................................... 713
  Rachit Arora and Balaraman Ravindran
INSCY: Indexing Subspace Clusters with In-Process-Removal of Redundancy ......................................................... 719
  Ira Assent, Ralph Krieger, Emmanuel Müller, and Thomas Seidl
A Conservative Feature Subset Selection Algorithm with Missing Data .................................................................................. 725
  Alex Aussem and Sergio Rodrigues de Morais
Nonparametric Monotone Classification with MOCA .............................................................................................................. 731
  Nicola Barile and Ad Feelders
Mining Large Networks with Subgraph Counting ........................................................................................................ 737
  Ilaria Bordino, Debora Donato, Aristides Gionis, and Stefano Leonardi
Comparative Evaluation of Anomaly Detection Techniques for Sequence Data ..................................................... 743
  Varun Chandola, Varun Mithal, and Vipin Kumar
On Locally Linear Classification by Pairwise Coupling ........................................................................................................ 749
  Feng Chen, Chang-Tien Lu, and Arnold P. Boedihardjo
A Probability Model for Projective Clustering on High Dimensional Data ........................................................................... 755
  Lifei Chen, Qingshan Jiang, and Shengrui Wang
Estimating Aggregates over Multiple Sets .................................................................761
Edith Cohen and Haim Kaplan

A Joint Matrix Factorization Approach to Unsupervised Action Categorization ........767
Peng Cui, Fei Wang, Li-Feng Sun, and Shi-Qiang Yang

Finding Alternative Clusterings Using Constraints ....................................................773
Ian Davidson and Zijie Qi

Efficient Feature Selection in the Presence of Multiple Feature Classes ....................779
Paramveer S. Dhillon, Dean Foster, and Lyle H. Ungar

Why Stacked Models Perform Effective Collective Classification ..........................785
Andrew Fast and David Jensen

Multiplicative Mixture Models for Overlapping Clustering .......................................791
Qiang Fu and Arindam Banerjee

Anomaly Detection Support Vector Machine and Its Application to Fault Diagnosis ....797
Ryohei Fujimaki

A Recommendation System for Preconditioned Iterative Solvers .............................803
Thomas George, Anshul Gupta, and Vivek Sarin

Cost-Sensitive Parsimonious Linear Regression .......................................................809
Robby Goetschaleckx, Kurt Driessens, and Scott Sanner

Text Mining in Radiology Reports ............................................................................815
Tianxia Gong, Chew Lim Tan, Tze Yun Leong, Cheng Kiang Lee, Boon Chuan Pang,
C. C. Tchoyoson Lim, Qi Tian, Suisheng Tang, and Zhuo Zhang

A Hierarchical Algorithm for Clustering Uncertain Data via an Information-Theoretic
Approach ..................................................................................................................821
Francesco Gullo, Giovanni Ponti, Andrea Tagarelli, and Sergio Greco

Discovering Significant Patterns in Multi-stream Sequences ....................................827
Robert Gwadera and Fabio Crestani

Graph-Based Rare Category Detection ......................................................................833
Jingrui He, Yan Liu, and Richard Lawrence

Clustering Documents with Active Learning Using Wikipedia ..................................839
Anna Huang, David Milne, Eibe Frank, and Ian H. Witten

Direct Zero-Norm Optimization for Feature Selection .............................................845
Kai Zhu Huang, Irwin King, and Michael R. Lyu

Discovering Flow Anomalies: A SWEET Approach ................................................851
James M. Kang, Shashi Shekhar, Christine Wennen, and Paige Novak

Boosting Relational Sequence Alignments ................................................................857
Andreas Karwath, Kristian Kersting, and Niels Landwehr

Support Vector Regression for Censored Data (SVRc): A Novel Tool for Survival
Analysis ..................................................................................................................863
Faisal M. Khan and Valentina Bayer Zubek

Nearest Neighbor Classifiers for Streaming Data with Delayed Labelling .................869
Ludmila I. Kuncheva and J. Salvador Sánchez

WiFIsViz: Effective Visualization of Frequent Itemsets ..........................................875
Carson Kai-Sang Leung, Pourang P. Irani, and Christopher L. Carmichael

x
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast and Memory Efficient Mining of High Utility Itemsets in Data Streams</td>
<td>881</td>
</tr>
<tr>
<td>Hua-Fu Li, Hsin-Yun Huang, Yi-Cheng Chen, Yu-Jiun Liu, and Suh-Yin Lee</td>
<td></td>
</tr>
<tr>
<td>HIREL: An Incremental Clustering Algorithm for Relational Datasets</td>
<td>877</td>
</tr>
<tr>
<td>Tao Li and Sarabjot S. Anand</td>
<td></td>
</tr>
<tr>
<td>Time Sensitive Ranking with Application to Publication Search</td>
<td>893</td>
</tr>
<tr>
<td>Xin Li, Bing Liu, and Philip Yu</td>
<td></td>
</tr>
<tr>
<td>Releasing the SVM Classifier with Privacy-Preservation</td>
<td>899</td>
</tr>
<tr>
<td>Keng-Pei Lin and Ming-Syan Chen</td>
<td></td>
</tr>
<tr>
<td>Text Cube: Computing IR Measures for Multidimensional Text Database Analysis</td>
<td>905</td>
</tr>
<tr>
<td>Cindy Xide Lin, Bolin Ding, Jiawei Han, Feida Zhu, and Bo Zhao</td>
<td></td>
</tr>
<tr>
<td>Multi-Space-Mapped SVMs for Multi-class Classification</td>
<td>911</td>
</tr>
<tr>
<td>Bo Liu, Longbing Cao, Philip S. Yu, and Chengqi Zhang</td>
<td></td>
</tr>
<tr>
<td>Spotting Significant Changing Subgraphs in Evolving Graphs</td>
<td>917</td>
</tr>
<tr>
<td>Zheng Liu, Jeffrey Xu Yu, Yiping Ke, Xuemin Lin, and Lei Chen</td>
<td></td>
</tr>
<tr>
<td>Classifying High-Dimensional Text and Web Data Using Very Short Patterns</td>
<td>923</td>
</tr>
<tr>
<td>Hassan H. Malik and John R. Kender</td>
<td></td>
</tr>
<tr>
<td>A Practical Approach to Classify Evolving Data Streams: Training with Limited Amount of Labeled Data</td>
<td>929</td>
</tr>
<tr>
<td>Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham</td>
<td></td>
</tr>
<tr>
<td>Spatiotemporal Relational Probability Trees: An Introduction</td>
<td>935</td>
</tr>
<tr>
<td>Amy McGovern, Nathan C. Hiers, Matthew Collier, David J. Gagne II, and Rodger A. Brown</td>
<td></td>
</tr>
<tr>
<td>Stream Sequential Pattern Mining with Precise Error Bounds</td>
<td>941</td>
</tr>
<tr>
<td>Luiz F. Mendes, Bolin Ding, and Jiawei Han</td>
<td></td>
</tr>
<tr>
<td>Organic Pie Charts</td>
<td>947</td>
</tr>
<tr>
<td>Fabian Moerchen</td>
<td></td>
</tr>
<tr>
<td>Frequent Subgraph Retrieval in Geometric Graph Databases</td>
<td>953</td>
</tr>
<tr>
<td>Sebastian Nowozin and Koji Tsuda</td>
<td></td>
</tr>
<tr>
<td>Alert Detection in System Logs</td>
<td>959</td>
</tr>
<tr>
<td>Adam J. Oliner, Alex Aiken, and Jon Stearley</td>
<td></td>
</tr>
<tr>
<td>Variance Minimization Least Squares Support Vector Machines for Time Series Analysis</td>
<td>965</td>
</tr>
<tr>
<td>Róbert Ormándi</td>
<td></td>
</tr>
<tr>
<td>Quantitative Association Analysis Using Tree Hierarchies</td>
<td>971</td>
</tr>
<tr>
<td>Feng Pan, Lynda Yang, Leonard McMillan, Fernando Pardo Manuel de Villena, David Threadgill, and Wei Wang</td>
<td></td>
</tr>
<tr>
<td>Sparse Maximum Margin Logistic Regression for Credit Scoring</td>
<td>977</td>
</tr>
<tr>
<td>Sabyasachi Patra, Kripa Shanker, and Debasis Kundu</td>
<td></td>
</tr>
<tr>
<td>Similarity Learning for Nearest Neighbor Classification</td>
<td>983</td>
</tr>
<tr>
<td>Ali Mustafa Qamar, Eric Gaussier, Jean-Pierre Chevallet, and Joo Hwee Lim</td>
<td></td>
</tr>
<tr>
<td>RBNBC: Repeat Based Naive Bayes Classifier for Biological Sequences</td>
<td>989</td>
</tr>
<tr>
<td>Pratibha Rani and Vikram Pudi</td>
<td></td>
</tr>
</tbody>
</table>
Multi-label Classification Using Ensembles of Pruned Sets
Jesse Read, Bernhard Pfahringer, and Geoff Holmes

Active Learning of Equivalence Relations by Minimizing the Expected Loss Using Constraint Inference
Steffen Rendle and Lars Schmidt-Thieme

Iterative Subgraph Mining for Principal Component Analysis
Hiroto Saigo and Koji Tsuda

Clustering Geospatial Objects via Hidden Markov Random Fields
Makoto Sato and Shuuichiro Imahara

Collective Latent Dirichlet Allocation
Zhi-Yong Shen, Jun Sun, and Yi-Dong Shen

Document-Word Co-regularization for Semi-supervised Sentiment Analysis
Vikas Sindhwani and Prem Melville

A Non-parametric Approach to Pair-Wise Dynamic Topic Correlation Detection
Yang Song, Lu Zhang, and C. Lee Giles

Block-Iterative Algorithms for Non-negative Matrix Approximation
Suvrit Sra

A Novel Method of Combined Feature Extraction for Recognition
Tingkai Sun, Songcan Chen, Jingyu Yang, and Pengfei Shi

Prediction of Skin Penetration Using Machine Learning Methods
Yi Sun, Gary P. Moss, Maria Prapopoulou, Rod Adams, Marc B. Brown, and Neil Davey

A Topic Modeling Approach and Its Integration into the Random Walk Framework for Academic Search
Jie Tang, Ruoming Jin, and Jing Zhang

Sequence Mining Automata: A New Technique for Mining Frequent Sequences under Regular Expressions
Roberto Trasarti, Francesco Bonchi, and Bart Goethals

Filling in the Blanks - Krimp Minimisation for Missing Data
Jilles Vreeken and Arno Siebes

Computational Discovery of Motifs Using Hierarchical Clustering Techniques
Dianhui Wang and Nung Kiong Lee

Inference Analysis in Privacy-Preserving Data Re-publishing
Guan Wang, Zutao Zhu, Wenliang Du, and Zhouxuan Teng

Using Wikipedia for Co-clustering Based Cross-Domain Text Classification
Pu Wang, Carlotta Domeniconi, and Jian Hu

Iterative Set Expansion of Named Entities Using the Web
Richard C. Wang and William W. Cohen

Experimental Evaluation of the Value of Structure: How to Efficiently Exploit Interdependencies in Sequence Labeling
Guillaume Wisniewski and Patrick Gallinari

Pseudolikelihood EM for Within-network Relational Learning
Rongjing Xiang and Jennifer Neville
Publishing Sensitive Transactions for Itemset Utility .................................................................1109
  Yabo Xu, Benjamin C. M. Fung, Ke Wang, Ada W. C. Fu, and Jian Pei
Learning the Latent Semantic Space for Ranking in Text Retrieval ........................................1115
  Jun Yan, Shuicheng Yan, Ning Liu, and Zheng Chen
Robust Time-Referenced Segmentation of Moving Object Trajectories ....................................1121
  Hyunjin Yoon and Cyrus Shahabi
Maximum Margin Embedding ....................................................................................................1127
  Bin Zhao, Fei Wang, and Changshui Zhang
Graph-Based Iterative Hybrid Feature Selection ......................................................................1133
  ErHeng Zhong, Sihong Xie, Wei Fan, Jiangtao Ren, Jing Peng, and Kun Zhang
Cleansing Noisy Data Streams ..................................................................................................1139
  Xingquan Zhu, Peng Zhang, Xindong Wu, Dan He, Chengqi Zhang, and Yong Shi

Author Index ................................................................................................................................1145