TABLE OF CONTENTS

ICDM’ 06: INVITED PAPERS

Neuroscience: New Insights for AI? .......................................................... 1
Tomaso Poggio

Exploratory Mining in Cube Space ......................................................... 4
Raghu Ramakrishnan

Data Mining Methods for Modeling Gene Expression Regulation and their Applications .......... 5
Weixiong Zhang

ICDM’ 06: REGULAR PAPERS

An Information Theoretic Approach to Detection of Minority Subsets in Database .................. 6
Shin Ando, Einoshin Suzuki

Bayesian State Space Modeling Approach for Measuring the Effectiveness of Marketing Activities and Baseline Sales from POS Data .................................................. 16
Tomohiro Ando

Learning to Use a Learned Model: A Two-Stage Approach to Classification ....................... 28
Maria-Luiza Antonie, Osmar Zaïane, Robert Holte

Hierarchical Classification by Expected Utility Maximization ........................................... 38
Korinna Bade, Eyke Hüllermeier, Andreas Nürnberger

COALA: A Novel Approach for the Extraction of an Alternate Clustering of High Quality and High Dissimilarity .......................................................... 48
Eric Bae, James Bailey

Cluster Ranking with an Application to Mining Mailbox Networks ..................................... 58
Ziv Bar-Yossef, Ido Guy, Ronny Lempel, Yoëlle Maarek, Vladimir Soroka

Large Scale Detection of Irregularities in Accounting Data ............................................. 70
Stephen Bay, Krishna Kumaraswamy, Markus Anderle, Rohit Kumar, David Steier

Adaptive Blocking: Learning to Scale Up Record Linkage ............................................... 82
Mikhail Bilenko, Beena Kamath, Raymond Mooney

Adaptive Parallel Graph Mining for CMP Architectures ................................................... 92
Gregory Buehrer, Srinivasan Parthasarathy, Yen-Kuang Chen

Meta Clustering ..................................................................................... 102
Rich Caruana, Mohamed Elhawary, Nam Nguyen, Casey Smith

Mixed-Drove Spatio-Temporal Co-occurrence Pattern Mining: A Summary of Results ............ 114
Mete Celik, Shashi Shekhar, James Rogers, James Shine, Jin Soung Yoo

An Interactive Semantic Video Mining and Retrieval Platform – Application in Transportation Surveillance Video for Incident Detection .............................................. 124
Xin Chen, Chengcui Zhang

Tolerance Closed Frequent Itemsets ............................................................................... 134
James Cheng, Yiping Ke, Wilfred Ng

Active Learning to Maximize Area Under the ROC Curve .............................................. 144
Matt Culver, Deng Kun, Stephen Scott
Integrating Features from Different Sources for Music Information Retrieval

Tao Li, Mitsunori Ogihara, Shenghuo Zhu

How Bayesians Debug

Chao Liu, Zeng Lian, Jiawei Han

On the Use of Structure and Sequence-Based Features for Protein Classification and Retrieval

Keith Marsolo, Srinivasan Parthasarathy

Accelerating Newton Optimization for Log-Linear Models through Feature Redundancy

Arpit Mathur, Soumen Chakrabarti

P3C: A Robust Projected Clustering Algorithm

Gabriela Moise, Jörg Sander, Martin Ester

Frequent Closed Itemset Mining Using Prefix Graphs with an Efficient Flow-based Pruning Strategy

H. Moonesinghe, Samah Fodeh, Pang-Ning Tan

Efficient Clustering of Uncertain Data

Wang Ngai, Ben Kao, Chun Chui, Reynold Cheng, Michael Chau, Kevin Yip

A Data Mining Approach for Capacity Building of Stakeholders in Integrated Flood Management

Peter Owotoki, Nataša Manojlovic, Friedrich Mayer-Lindenberg, Erik Pasche

Local Correlation Tracking in Time Series

Spiros Papadimitriou, Jimeng Sun, Philip Yu

Who Thinks Who Knows Who? Socio-cognitive Analysis of Email Networks

Nishith Pathak, Sandeep Mane, Jaideep Srivastava

An Efficient Reference-based Approach to Outlier Detection in Large Datasets

Yaling Pei, Osmar Zaïane, Yong Gao

Using an Ensemble of One-Class SVM Classifiers to Harden Payload-based Anomaly Detection Systems

Roberto Perdisci, Guofei Gu, Wenke Lee

Relational Ensemble Classification

Christine Preisach, Lars Schmidt-Thieme

Discovering Partial Orders in Binary Data

Deepak Rajan, Philip Yu

Stability Region Based Expectation Maximization for Model-based Clustering

Chandan Reddy, Hsiao-Dong Chiang, Bala Rajaratnam

Co-clustering Documents and Words Using Bipartite Isoperimetric Graph Partitioning

Manjeet Rege, Ming Dong, Farshad Fotouhi

Latent Dirichlet Co-Clustering

M. Shafiei, Evangelos Milios

Latent Friend Mining from Blog Data

Dou Shen, Jian-Tao Sun, Qiang Yang, Zheng Chen

The PDD Framework for Detecting Categories of Peculiar Data

Mahesh Shrestha, Howard Hamilton, Yiyu Yao, Ken Konkel, Liqiang Geng

Entity Resolution with Markov Logic

Parag Singla, Pedro Domingos
Boosting Kernel Models for Regression ........................................................................................................578
Ping Sun, Xin Yao

Boosting for Learning Multiple Classes with Imbalanced Class Distribution ........................................587
Yanmin Sun, Mohamed Kamel, Yang Wang

What is the Dimension of your Binary Data? ...............................................................................................598
Nikolaj Tatti, Taneli Mielikäinen, Aristides Gionis, Heikki Mannila

Fast Random Walk with Restart and its Applications .............................................................................608
Hanghang Tong, Christos Faloutsos, Jia-Yu Pan

Anytime Classification Using the Nearest Neighbor Algorithm with Applications to Stream Mining ........618
Ken Ueno, Xiaopeng Xi, Eamonn Keogh, Dah-Jye Lee

Dirichlet Aspect Weighting: A Generalized EM Algorithm for Integrating External Data Fields with Semantically Structured Queries by Using Gradient Projection Method ....................628
Atulya Velivelli, Thomas Huang

Lazy Associative Classification ..................................................................................................................640
Adriano Veloso, Wagner Meira Jr., Mohammed Zakib

Geometrically Inspired Itemset Mining ........................................................................................................650
Florian Verhein, Sanjay Chawla

Finding “Who is Talking to Whom” in VoIP Networks via Progressive Stream Clustering ...................662
Olivier Verscheure, Michail Vlachos, Aris Anagnostopoulos, Pascal Frossard, Eric Bouillet, Philip Yu

Comparison of Descriptor Spaces for Chemical Compound Retrieval and Classification ..........................673
Nikil Wale, George Karypis

Regularized Least Absolute Deviations Regression and an Efficient Algorithm for Parameter Tuning ....685
Li Wang, Michael Gordon, Ji Zhu

LOC1: Load Shedding through Class-Preserving Data Acquisition .........................................................696
Peng Wang, Haixun Wang, Wei Wang, Baile Shi, Philip Yu

SAXually Explicit Images: Finding Unusual Shapes ...................................................................................706
Li Wei, Eamonn Keogh, Xiaopeng Xi

A Novel Scalable Algorithm for Supervised Subspace Learning ...............................................................716
Jun Yan, Ning Liu, Benyu Zhang, Qiang Yang, Shuicheng Yan, Zheng Chen

A Novel Method for Detecting Outlying Subspaces in High-dimensional Databases Using Genetic Algorithm ..........................................................................................................................726
Ji Zhang, Qiqiang Gao, Hai Wang

Discovering Unrevealed Properties of Probability Estimation Trees: On Algorithm Selection and Performance Explanation .............................................................................................................736
Kun Zhang, Wei Fan, Bill Buckles, Xiaojing Yuan, Zujia Xu

Forecasting Skewed Biased Stochastic Ozone Days: Analyses and Solutions ........................................748
Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, Xiangshang Li

Identifying Follow-Correlation Itemset-Pairs ...............................................................................................760
Shichao Zhang, Jilian Zhang, Xiaofeng Zhu, Zifang Huang

On the Lower Bound of Local Optimums in K-Means Algorithm ...............................................................770
Zhenjie Zhang, Bing Tian Dai, Anthony Tang
Fast On-line Kernel Learning for Trees ............................................................. 780

Fabio Aiolli, Giovanni Da San Martino, Alessandro Sperduti, Alessandro Moschitti

bitSPADE: A Lattice-based Sequential Pattern Mining Algorithm Using Bitmap Representation ................................................................. 785

Sujeevan Aseervatham, Aomar Osmani, Emmanuel Viennet

Decision Trees for Functional Variables .......................................................... 791

Suhrid Balakrishnan, David Madigan

Mining Latent Associations of Objects Using a Typed Mixture Model — A Case Study on Expert/Expertise Mining .................................................. 796

Shenghua Bao, Yunbo Cao, Bing Liu, Yong Yu, Hang Li

Semantic Kernels for Text Classification Based on Topological Measures of Feature Similarity ................................................................................. 801

Stephan Bloehdorn, Roberto Basili, Marco Cammisa, Alessandro Moschitti

Mining Maximal Generalized Frequent Geographic Patterns with Knowledge Constraints .............................................................. 806

Vania Bogorny, João Valiati, Sandro Camargo, Paulo Engel, Bart Kuipers, Luis Alves

Pattern Mining in Frequent Dynamic Subgraphs ........................................... 811

Karsten Borgwardt, Hans-Peter Kriegel, Peter Wackersreuther

Discovery of Collocation Episodes in Spatiotemporal Data .............................. 816

Huiping Cao, Nikos Mamoulis, David Cheung

Getting the Most Out of Ensemble Selection .................................................. 821

Rich Caruana, Art Munson, Alexandru Niculescu-Mizil

Diverse Topic Phrase Extraction through Latent Semantic Analysis .................... 827

Jilin Chen, Jun Yan, Benyu Zhang, Qiang Yang, Zheng Chen

AC-Close: Efficiently Mining Approximate Closed Itemsets by Core Pattern Recovery ................................................................. 832

Hong Cheng, Philip Yu, Jiawei Han

Belief Propagation in Large, Highly Connected Graphs for 3D Part-Based Object Recognition ................................................................. 838

Frank DiMaio, Jude Shavlik

A Framework for Regional Association Rule Mining in Spatial Datasets .............. 844

Wei Ding, Christoph Eick, Jing Wang, Xiaojing Yuan

Mining Generalized Graph Patterns Based on User Examples ............................ 850

Pavel Dmitriev, Carl Lagoze

An Experimental Investigation of Graph Kernels on a Collaborative Recommendation Task ...................................................................................... 856

François Fouss, Luh Yen, Alain Pirrotte, Marco Saerens

A Balanced Ensemble Approach to Weighting Classifiers for Text Classification ..................................................................................... 862

Gabriel Fung, Jeffrey Yu, Haixun Wang, David Cheung, Huan Liu

Detection of Interdomain Routing Anomalies Based on Higher-Order Path Analysis ................................................................. 867

Murat Can Gantz, Sudhan Kantikar, Mooi Choo Chuah, William Pottenger

Star-Structured High-Order Heterogeneous Data Co-clustering Based on Consistent Information Theory ............................................................. 873

Bin Gao, Tie-Yan Liu, Wei-Ying Ma
GraphRank: Statistical Modeling and Mining of Significant Subgraphs in the Feature Space ................................................................. 878
Huahai He, Ambuj Singh

A Feature Selection and Evaluation Scheme for Computer Virus Detection .................................................................................. 884
Olivier Henchiri, Nathalie Japkowicz

Cluster Analysis of Time-series Medical Data Based on the Trajectory Representation and Multiscale Comparison Techniques .......................................................... 889
Shoji Hirano, Shusaku Tsumoto

Constructing Ensembles for Better Ranking ................................................................................................................................. 895
Jin Huang, Charles X. Ling

TRIAS — An Algorithm for Mining Iceberg Tri-Lattices .................................................................................................................. 900
Robert Jäschke, Andreas Hotho, Christoph Schmitz, Bernhard Ganter, Gerd Stumme

Intelligent Icons: Integrating Lite-Weight Data Mining and Visualization into GUI Operating Systems .................................................................................................................. 905
Eamonn Keogh, Li Wei, Xiaopeng Xi, Stefano Lonardi, Jin Shieh, Scott Sirowy

COSMIC: Conceptually Specified Multi-Instance Clusters .......................................................................................................... 910
Hans-Peter Kriegel, Alexey Pryakhin, Matthias Schubert, Arthur Zimek

Direct Marketing When There Are Voluntary Buyers .................................................................................................................. 915
Yi-Ting Lai, Ke Wang, Daymond Ling, Hua Shi, Jason Zhang

DSTree: A Tree Structure for the Mining of Frequent Sets from Data Streams ........................................................................... 921
Carson Leung, Quamrul Khan

Searching for Pattern Rules ................................................................................................................................................................. 926
Guichong Li, Howard Hamilton

Adding Semantics to Email Clustering ............................................................................................................................................... 931
Hua Li, Dou Shen, Benyu Zhang, Zheng Chen, Qiang Yang

Gradual Cube: Customize Profile on Mobile OLAP ......................................................................................................................... 936
Jun Li, HaoFeng Zhou, Wei Wang

CoMiner: An Effective Algorithm for Mining Competitors from the Web .................................................................................. 941
Rui Li, Shenghua Bao, Jin Wang, Yong Yu, Yunbo Cao

Multi-Tier Granule Mining for Representations of Multidimensional Association Rules .................................................................................. 946
Yuefeng Li, Wanzhong Yang, Yue Xu

Social Capital in Friendship-Event Networks ........................................................................................................................................ 952
Louis Licamele, Lise Getoor

Exploratory Under-Sampling for Class-Imbalance Learning ........................................................................................................ 958
Xu-Ying Liu, Jianxin Wu, Zhi-Hua Zhou

The Influence of Class Imbalance on Cost-Sensitive Learning: An Empirical Study ........................................................................ 963
Xu-Ying Liu, Zhi-Hua Zhou

Similarity of Temporal Query Logs Based on ARIMA Model ...................................................................................................... 968
Ning Liu, Shuzhen Nong, Jun Yan, Benyu Zhang, Zheng Chen, Ying Li

Probabilistic Segmentation and Analysis of Horizontal Cells .................................................................................................... 973
Vebjorn Ljosa, Ambuj Singh

Mining Correlation between Motifs and Gene Expression .......................................................................................................... 979
Yi Lu, Shiyong Lu, Adrian Platts, Stephen Krawetz
High Quality, Efficient Hierarchical Document Clustering Using Closed Interesting Itemsets
Hassan Malik, John Kender

On Trajectory Representation for Scientific Features
Sameep Mehta, Srinivasan Parthasarathy, Raghu Machiraju

NewsCATS: A News Categorization and Trading System
Marc-André Mittermayer, Gerhard Knollmayer

Improving Grouped-Entity Resolution Using Quasi-Cliques
Byung-Won On, Ergin Elmacioglu, Dongwon Lee, Jaewoo Kang, Jian Pei

Fast Relevance Discovery in Time Series
Chang-shing Perng, Haixun Wang, Sheng Ma

Probabilistic Enhanced Mapping with the Generative Tabular Model
Rodolphe Priam, Mohamed Nadif

Object Identification with Constraints
Steffen Rendle, Lars Schmidt-Thieme

High-Performance Unsupervised Relation Extraction from Large Corpora
Binjamin Rozenfeld, Ronen Feldman

Cluster Based Core Vector Machine
Asharaf S., M. Murty, S. Shevade

Enhancing Text Clustering Using Concept-based Mining Model
Shady Shehata, Fakhri Karray, Mohamed Kamel

Detecting Link Spam using Temporal Information
Guoyang Shen, Bin Gao, Tie-Yan Liu, Guang Feng, Shiji Song, Hang Li

Minimum Enclosing Spheres Formulations for Support Vector Ordinal Regression
S. Shevade, Wei Chu

Mining Maximal Quasi-Bicliques to Co-Cluster Stocks and Financial Ratios for Value Investment
Kelvin Sim, Jinyan Li, Vivekanand Gopalkrishnan, Guimei Liu

Boosting the Feature Space: Text Classification for Unstructured Data on the Web
Yang Song, Ding Zhou, Jian Huang, Isaac Councill, Hongyuan Zha, C. Giles

Plagiarism Detection in arXiv
Daria Sorokina, Johannes Gehrke, Simeon Warner, Paul Ginsparg

Window-based Tensor Analysis on High-dimensional and Multi-aspect Streams
Jimeng Sun, Spiros Papadimitriou, Philip Yu

Automatic Single-Organ Segmentation in Computed Tomography Images
Ruchaneewan Susomboon, Daniela Raicu, Jacob Furst, David Channin

Improving Nearest Neighbor Classifier Using Tabu Search and Ensemble Distance Metrics
Muhammad Tahir, James Smith

Comparisons of K-Anonymization and Randomization Schemes Under Linking Attacks
Zhouxuan Teng, Wenliang Du

MARGIN: Maximal Frequent Subgraph Mining
Lini Thomas, Satyanarayana Valluri, Kamalakar Karlapalem

Resource Management for Networked Classifiers in Distributed Stream Mining Systems
Deepak Turaga, Olivier Verscheure, Upenbra Chaudhari, Lisa Amini
A Simple Yet Effective Data Clustering Algorithm .................................................................1101
Soujanya Vadapalli, Satyanarayana Valluri, Kamalakar Karlapalem

Entropy-based Concept Shift Detection ..............................................................................1106
Peter Vorburger, Abraham Bernstein

Recommendation on Item Graphs ....................................................................................1112
Fei Wang, Sheng Ma, Lichong Yang, Tao Li

Solution Path for Semi-Supervised Classification with Manifold Regularization ..........1117
Gang Wang, Tao Chen, Dit-Yan Yeung, Frederick Lochovksy

Semi-Supervised Kernel Regression ...................................................................................1123
Meng Wang, Xian-Sheng Hua, Yan Song, Li-Rong Dai, Hong-Jiang Zhang

Mining Complex Time-Series Data by Learning Markovian Models .........................1129
Yi Wang, Lizhu Zhou, Jianhua Feng, Jianyong Wang, Zhi-Qiang Liu

Temporal Data Mining in Dynamic Feature Spaces ........................................................1134
Brent Wenerstrom, Christophe Giraud-Carrier

Discover Bayesian Networks from Incomplete Data Using a Hybrid Evolutionary
Algorithm .................................................................................................................................1139
Man Leung Wong, Yuan Yuan Guo

Distances and (Indefinite) Kernels for Sets of Objects ....................................................1144
Adam Woznica, Alexandros Kalousis, Melanie Hilario

Deploying Approaches for Pattern Refinement in Text Mining ........................................1150
Sheng-Tang Wu, Yuefeng Li, Yue Xu

TOP-COP: Mining TOP-K Strongly Correlated Pairs in Large Databases ...................1155
Hui Xiong, Mark Brodie, Sheng Ma

Manifold Clustering of Shapes .........................................................................................1160
Dragomir Yankov, Eamonn Keogh

Linear and Non-Linear Dimensional Reduction via Class Representatives for Text
Classification ............................................................................................................................1165
Dimitrios Zeimpekis, Efstratios Gallopoulos

Adaptive Kernel Principal Component Analysis with Unsupervised Learning of Kernels 1171
Daoqiang Zhang, Zhi-Hua Zhou, Songcan Chen

Rule-Based Platform for Web User Profiling ..................................................................1176
Jianping Zhang, Manu Shukla

Opening the Black Box of Feature Extraction: Incorporating Visualization into High-
Dimensional Data Mining Processes .................................................................................1181
Jianting Zhang, Le Gruenwald

Semantic Smoothing for Model-based Document Clustering ........................................1186
Xiaodan Zhang, Xiaohua Zhou, Xiaohua Hu

Corrective Classification: Classifier Ensembling with Corrective and Diverse Base
Learners .................................................................................................................................1192
Yan Zhang, Xingsuan Zhu, Xindong Wu

Speedup Clustering with Hierarchical Ranking ................................................................1198
Jianjun Zhou, Joerg Sander

Query-Sensitive Similarity Measure for Content-Based Image Retrieval .....................1204
Zhi-Hua Zhou, Hong-Bin Dai

Author Index