

# Table of Contents – Part I

## Keynote Speeches

Empower People with Knowledge: The Next Frontier for Web Search ... <i>Wei-Ying Ma</i>	1
Discovery of Patterns in Global Earth Science Data Using Data Mining ..... <i>Vipin Kumar</i>	2
Game Theoretic Approaches to Knowledge Discovery and Data Mining ..... <i>Y. Narahari</i>	3

## Session 1A. Clustering I

A Set Correlation Model for Partitional Clustering ..... <i>Nguyen Xuan Vinh and Michael E. Houle</i>	4
iVAT and aVAT: Enhanced Visual Analysis for Cluster Tendency Assessment ..... <i>Liang Wang, Uyen T.V. Nguyen, James C. Bezdek, Christopher A. Leckie, and Kotagiri Ramamohanarao</i>	16
A Robust Seedless Algorithm for Correlation Clustering ..... <i>Mohammad S. Aziz and Chandan K. Reddy</i>	28
Integrative Parameter-Free Clustering of Data with Mixed Type Attributes ..... <i>Christian Böhm, Sebastian Goebel, Annahita Oswald, Claudia Plant, Michael Plavinski, and Bianca Wackersreuther</i>	38
Data Transformation for Sum Squared Residue ..... <i>Hyuk Cho</i>	48

## Session 1B. Social Networks

A Better Strategy of Discovering Link-Pattern Based Communities by Classical Clustering Methods ..... <i>Chen-Yi Lin, Jia-Ling Koh, and Arbee L.P. Chen</i>	56
Mining Antagonistic Communities from Social Networks ..... <i>Kuan Zhang, David Lo, and Ee-Peng Lim</i>	68

As Time Goes by: Discovering Eras in Evolving Social Networks . . . . .	81
<i>Michele Berlingiero, Michele Coscia, Fosca Giannotti, Anna Monreale, and Dino Pedreschi</i>	
Online Sampling of High Centrality Individuals in Social Networks . . . . .	91
<i>Arun S. Maiya and Tanya Y. Berger-Wolf</i>	
Estimate on Expectation for Influence Maximization in Social Networks . . . . .	99
<i>Yao Zhang, Qing Gu, Jun Zheng, and Daoxu Chen</i>	

**Session 1C. Classification I**

A Novel Scalable Multi-class ROC for Effective Visualization and Computation . . . . .	107
<i>Md. Rafiul Hassan, Kotagiri Ramamohanarao, Chandan Karmakar, M. Maruf Hossain, and James Bailey</i>	
Efficiently Finding the Best Parameter for the Emerging Pattern-Based Classifier PCL . . . . .	121
<i>Thanh-Son Ngo, Mengling Feng, Guimei Liu, and Limsoon Wong</i>	
Rough Margin Based Core Vector Machine . . . . .	134
<i>Gang Niu, Bo Dai, Lin Shang, and Yangsheng Ji</i>	
BoostML: An Adaptive Metric Learning for Nearest Neighbor Classification . . . . .	142
<i>Nayyar Abbas Zaidi, David McG. Squire, and David Suter</i>	
A New Emerging Pattern Mining Algorithm and Its Application in Supervised Classification . . . . .	150
<i>Milton García-Borroto, José Francisco Martínez-Trinidad, and Jesús Ariel Carrasco-Ochoa</i>	

**Session 2A. Privacy**

Hiding Emerging Patterns with Local Recoding Generalization . . . . .	158
<i>Michael W.K. Cheng, Byron Koon Kau Choi, and William Kwok Wai Cheung</i>	
Anonymizing Transaction Data by Integrating Suppression and Generalization . . . . .	171
<i>Junqiang Liu and Ke Wang</i>	
Satisfying Privacy Requirements: One Step before Anonymization . . . . .	181
<i>Xiaoxun Sun, Hua Wang, and Jiuyong Li</i>	

Computation of Ratios of Secure Summations in Multi-party Privacy-Preserving Latent Dirichlet Allocation .....	189
<i>Bin Yang and Hiroshi Nakagawa</i>	
Privacy-Preserving Network Aggregation .....	198
<i>Troy Raeder, Marina Blanton, Nitesh V. Chawla, and Keith Frikken</i>	
Multivariate Equi-width Data Swapping for Private Data Publication...	208
<i>Yidong Li and Hong Shen</i>	

## Session 2B. Spatio-Temporal Mining

Correspondence Clustering: An Approach to Cluster Multiple Related Spatial Datasets .....	216
<i>Vadeerat Rinsurongkawong and Christoph F. Eick</i>	
Mining Trajectory Corridors Using Fréchet Distance and Meshing Grids .....	228
<i>Haohan Zhu, Jun Luo, Hang Yin, Xiaotao Zhou, Joshua Zherue Huang, and F. Benjamin Zhan</i>	
Subseries Join: A Similarity-Based Time Series Match Approach .....	238
<i>Yi Lin and Michael D. McCool</i>	
TWave: High-Order Analysis of Spatiotemporal Data .....	246
<i>Michael Barnathan, Vasileios Megalooikonomou, Christos Faloutsos, Feroze B. Mohamed, and Scott Faro</i>	
Spatial Clustering with Obstacles Constraints by Dynamic Piecewise-Mapped and Nonlinear Inertia Weights PSO .....	254
<i>Xueping Zhang, Haohua Du, and Jiayao Wang</i>	

## Session 3A. Pattern Mining

An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns .....	262
<i>Zhigang Zheng, Yanchang Zhao, Ziye Zuo, and Longbing Cao</i>	
Valency Based Weighted Association Rule Mining .....	274
<i>Yun Sing Koh, Russel Pears, and Wai Yeap</i>	
Ranking Sequential Patterns with Respect to Significance .....	286
<i>Robert Gwadera and Fabio Crestani</i>	
Mining Association Rules in Long Sequences .....	300
<i>Boris Cule and Bart Goethals</i>	
Mining Closed Episodes from Event Sequences Efficiently .....	310
<i>Wenzhi Zhou, Hongyan Liu, and Hong Cheng</i>	

Most Significant Substring Mining Based on Chi-square Measure . . . . .	319
<i>Sourav Dutta and Arnab Bhattacharya</i>	

### Session 3B. Recommendations/Answers

Probabilistic User Modeling in the Presence of Drifting Concepts . . . . .	328
<i>Vikas Bhardwaj and Ramaswamy Devarajan</i>	
Using Association Rules to Solve the Cold-Start Problem in Recommender Systems . . . . .	340
<i>Gavin Shaw, Yue Xu, and Shlomo Geva</i>	
Semi-supervised Tag Recommendation - Using Untagged Resources to Mitigate Cold-Start Problems . . . . .	348
<i>Christine Preisach, Leandro Balby Marinho, and Lars Schmidt-Thieme</i>	
Cost-Sensitive Listwise Ranking Approach . . . . .	358
<i>Min Lu, MaoQiang Xie, Yang Wang, Jie Liu, and YaLou Huang</i>	
Mining Wikipedia and Yahoo! Answers for Question Expansion in Opinion QA . . . . .	367
<i>Yajie Miao and Chunping Li</i>	
Answer Diversification for Complex Question Answering on the Web ...	375
<i>Palakorn Achananuparp, Xiaohua Hu, Tingting He, Christopher C. Yang, Yuan An, and Lifan Guo</i>	
Vocabulary Filtering for Term Weighting in Archived Question Search . . . . .	383
<i>Zhao-Yan Ming, Kai Wang, and Tat-Seng Chua</i>	

### Session 3C. Topic Modeling/Information Extraction

On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations . . . . .	391
<i>R. Arun, V. Suresh, C.E. Veni Madhavan, and M.N. Narasimha Murthy</i>	
Supervising Latent Topic Model for Maximum-Margin Text Classification and Regression . . . . .	403
<i>Wanhong Xu</i>	
Resource-bounded Information Extraction: Acquiring Missing Feature Values On Demand . . . . .	415
<i>Pallika Kanani, Andrew McCallum, and Shaohan Hu</i>	
Efficient Deep Web Crawling Using Reinforcement Learning . . . . .	428
<i>Lu Jiang, Zhaohui Wu, Qian Feng, Jun Liu, and Qinghua Zheng</i>	

Topic Decomposition and Summarization .....	440
<i>Wei Chen, Can Wang, Chun Chen, Lijun Zhang, and Jiajun Bu</i>	
<b>Session 4A. Skylines/Uncertainty</b>	
UNN: A Neural Network for Uncertain Data Classification .....	449
<i>Jiaqi Ge, Yuni Xia, and Chandima Nadungodage</i>	
SkyDist: Data Mining on Skyline Objects .....	461
<i>Christian Böhm, Annahita Oswald, Claudia Plant, Michael Plavinski, and Bianca Wackersreuther</i>	
Multi-Source Skyline Queries Processing in Multi-Dimensional Space ...	471
<i>Cuiping Li, Wenlin He, and Hong Chen</i>	
Efficient Pattern Mining of Uncertain Data with Sampling .....	480
<i>Toon Calders, Calin Garboni, and Bart Goethals</i>	
Classifier Ensemble for Uncertain Data Stream Classification .....	488
<i>Shirui Pan, Kuan Wu, Yang Zhang, and Xue Li</i>	
<b>Author Index</b> .....	497

## Table of Contents – Part II

### Session 4B. Dimensionality Reduction/Parallelism

Subclass-Oriented Dimension Reduction with Constraint Transformation and Manifold Regularization .....	1
<i>Bin Tong and Einoshin Suzuki</i>	
Distributed Knowledge Discovery with Non Linear Dimensionality Reduction .....	14
<i>Panagis Magdalinos, Michalis Vazirgiannis, and Dialecti Valsamou</i>	
DPSP: Distributed Progressive Sequential Pattern Mining on the Cloud .....	27
<i>Jen-Wei Huang, Su-Chen Lin, and Ming-Syan Chen</i>	
An Approach for Fast Hierarchical Agglomerative Clustering Using Graphics Processors with CUDA .....	35
<i>S.A. Arul Shalom, Manoranjan Dash, and Minh Tue</i>	

### Session 5A. Novel Applications

Ontology-Based Mining of Brainwaves: A Sequence Similarity Technique for Mapping Alternative Features in Event-Related Potentials (ERP) Data .....	43
<i>Haishan Liu, Gwen Frishkoff, Robert Frank, and Dejing Dou</i>	
Combining Support Vector Machines and the $t$ -statistic for Gene Selection in DNA Microarray Data Analysis .....	55
<i>Tao Yang, Vojislave Kecman, Longbing Cao, and Chengqi Zhang</i>	
Satrap: Data and Network Heterogeneity Aware P2P Data-Mining .....	63
<i>Hock Hee Ang, Vivekanand Gopalkrishnan, Anwitaman Datta, Wee Keong Ng, and Steven C.H. Hoi</i>	
Player Performance Prediction in Massively Multiplayer Online Role-Playing Games (MMORPGs) .....	71
<i>Kyong Jin Shim, Richa Sharan, and Jaideep Srivastava</i>	
Relevant Gene Selection Using Normalized Cut Clustering with Maximal Compression Similarity Measure .....	81
<i>Rajni Bala, R.K. Agrawal, and Manju Sardana</i>	

**Session 5B. Feature Selection/Visualization**

A Novel Prototype Reduction Method for the $K$ -Nearest Neighbor Algorithm with $K \geq 1$ .....	89
<i>Tao Yang, Longbing Cao, and Chengqi Zhang</i>	
Generalized Two-Dimensional FLD Method for Feature Extraction: An Application to Face Recognition .....	101
<i>Shiladitya Chowdhury, Jamuna Kanta Sing, Dipak Kumar Basu, and Mita Nasipuri</i>	
Learning Gradients with Gaussian Processes.....	113
<i>Xinwei Jiang, Junbin Gao, Tianjiang Wang, and Paul W. Kwan</i>	
Analyzing the Role of Dimension Arrangement for Data Visualization in Radviz .....	125
<i>Luigi Di Caro, Vanessa Frias-Martinez, and Enrique Frias-Martinez</i>	

**Session 6A. Graph Mining**

Subgraph Mining on Directed and Weighted Graphs.....	133
<i>Stephan Günnemann and Thomas Seidl</i>	
Finding Itemset-Sharing Patterns in a Large Itemset-Associated Graph .....	147
<i>Mutsumi Fukuzaki, Mio Seki, Hisashi Kashima, and Jun Sese</i>	
A Framework for SQL-Based Mining of Large Graphs on Relational Databases.....	160
<i>Sriganesh Srihari, Shruti Chandrashekar, and Srinivasan Parthasarathy</i>	
Fast Discovery of Reliable $k$ -terminal Subgraphs .....	168
<i>Melissa Kasari, Hannu Toivonen, and Petteri Hintsanen</i>	
GTRACE2: Improving Performance Using Labeled Union Graphs .....	178
<i>Akihiro Inokuchi and Takashi Washio</i>	

**Session 6B. Clustering II**

Orthogonal Nonnegative Matrix Tri-factorization for Semi-supervised Document Co-clustering .....	189
<i>Huifang Ma, Weizhong Zhao, Qing Tan, and Zhongzhi Shi</i>	
Rule Synthesizing from Multiple Related Databases .....	201
<i>Dan He, Xindong Wu, and Xingquan Zhu</i>	

Fast Orthogonal Nonnegative Matrix Tri-Factorization for Simultaneous Clustering .....	214
<i>Zhao Li, Xindong Wu, and Zhenyu Lu</i>	

Hierarchical Web-Page Clustering via In-Page and Cross-Page Link Structures .....	222
<i>Cindy Xide Lin, Yintao Yu, Jiawei Han, and Bing Liu</i>	

Mining Numbers in Text Using Suffix Arrays and Clustering Based on Dirichlet Process Mixture Models .....	230
<i>Minoru Yoshida, Issei Sato, Hiroshi Nakagawa, and Akira Terada</i>	

## Session 7A. Opinion/Sentiment Mining

Opinion-Based Imprecise Query Answering .....	238
<i>Muhammad Abulaish, Tanvir Ahmad, Jahiruddin, and Mohammad Najmud Doja</i>	

Blog Opinion Retrieval Based on Topic-Opinion Mixture Model .....	249
<i>Peng Jiang, Chunxia Zhang, Qing Yang, and Zhendong Niu</i>	

Feature Subsumption for Sentiment Classification in Multiple Languages .....	261
<i>Zhongwu Zhai, Hua Xu, Jun Li, and Peifa Jia</i>	

Decentralisation of ScoreFinder: A Framework for Credibility Management on User-Generated Contents .....	272
<i>Yang Liao, Aaron Harwood, and Kotagiri Ramamohanarao</i>	

Classification and Pattern Discovery of Mood in Weblogs .....	283
<i>Thin Nguyen, Dinh Phung, Brett Adams, Truyen Tran, and Svetha Venkatesh</i>	

Capture of Evidence for Summarization: An Application of Enhanced Subjective Logic .....	291
<i>Sukanya Manna, B. Sumudu U. Mendis, and Tom Gedeon</i>	

## Session 7B. Stream Mining

Fast Perceptron Decision Tree Learning from Evolving Data Streams ...	299
<i>Albert Bifet, Geoff Holmes, Bernhard Pfahringer, and Eibe Frank</i>	

Classification and Novel Class Detection in Data Streams with Active Mining .....	311
<i>Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham</i>	

Bulk Loading Hierarchical Mixture Models for Efficient Stream  
Classification ..... 325  
*Philipp Kranen, Ralph Krieger, Stefan Denker, and Thomas Seidl*

Summarizing Multidimensional Data Streams: A Hierarchy-Graph-  
Based Approach ..... 335  
*Yoann Pitarch, Anne Laurent, and Pascal Poncelet*

Efficient Trade-Off between Speed Processing and Accuracy in  
Summarizing Data Streams ..... 343  
*Nesrine Gabsi, Fabrice Cl  rot, and Georges H  brail*

Subsequence Matching of Stream Synopses under the Time Warping  
Distance ..... 354  
*Su-Chen Lin, Mi-Yen Yeh, and Ming-Syan Chen*

**Session 8A. Similarity and Kernels**

Normalized Kernels as Similarity Indices ..... 362  
*Julien Ah-Pine*

Adaptive Matching Based Kernels for Labelled Graphs ..... 374  
*Adam Wo  znica, Alexandros Kalousis, and Melanie Hilario*

A New Framework for Dissimilarity and Similarity Learning ..... 386  
*Adam Wo  znica and Alexandros Kalousis*

Semantic-Distance Based Clustering for XML Keyword Search ..... 398  
*Weidong Yang and Hao Zhu*

**Session 8B. Graph Analysis**

OddBall: Spotting Anomalies in Weighted Graphs ..... 410  
*Leman Akoglu, Mary McGlohon, and Christos Faloutsos*

Robust Outlier Detection Using Commute Time and Eigenspace  
Embedding ..... 422  
*Nguyen Lu Dang Khoa and Sanjay Chawla*

EigenSpokes: Surprising Patterns and Scalable Community Chipping  
in Large Graphs ..... 435  
*B. Aditya Prakash, Ashwin Sridharan, Mukund Seshadri,  
Sridhar Machiraju, and Christos Faloutsos*

BASSET: Scalable Gateway Finder in Large Graphs ..... 449  
*Hanghang Tong, Spiros Papadimitriou, Christos Faloutsos,  
Philip S. Yu, and Tina Eliassi-Rad*

**Session 8C. Classification II**

Ensemble Learning Based on Multi-Task Class Labels .....	464
<i>Qing Wang and Liang Zhang</i>	
Supervised Learning with Minimal Effort .....	476
<i>Eileen A. Ni and Charles X. Ling</i>	
Generating Diverse Ensembles to Counter the Problem of Class Imbalance .....	488
<i>T. Ryan Hoens and Nitesh V. Chawla</i>	
Relationship between Diversity and Correlation in Multi-Classifer Systems .....	500
<i>Kuo-Wei Hsu and Jaideep Srivastava</i>	
Compact Margin Machine .....	507
<i>Bo Dai and Gang Niu</i>	
<b>Author Index</b> .....	515