

# Table of Contents

## Industrial Papers (Invited)

Data Mining as an Automated Service .....	1
<i>P.S. Bradley</i>	
Trends and Challenges in the Industrial Applications of KDD .....	14
<i>Ramasamy Uthurusamy</i>	

## Session 1A: Stream Mining I

Finding Event-Oriented Patterns in Long Temporal Sequences .....	15
<i>Xingzhi Sun, Maria E. Orlowska, Xiaofang Zhou</i>	
Mining Frequent Episodes for Relating Financial Events and Stock Trends .....	27
<i>Anny Ng, Ada Wai-chee Fu</i>	

## Session 1B: Graph Mining

An Efficient Algorithm of Frequent Connected Subgraph Extraction .....	40
<i>Mingsheng Hong, Haofeng Zhou, Wei Wang, Baile Shi</i>	
Classifier Construction by Graph-Based Induction for Graph-Structured Data .....	52
<i>Warodom Geamsakul, Takashi Matsuda, Tetsuya Yoshida, Hiroshi Motoda, Takashi Washio</i>	

## Session 1C: Clustering I

Comparison of the Performance of Center-Based Clustering Algorithms .....	63
<i>Bin Zhang</i>	
Automatic Extraction of Clusters from Hierarchical Clustering Representations .....	75
<i>Jörg Sander, Xuejie Qin, Zhiyong Lu, Nan Niu, Alex Kovarsky</i>	

## Session 2A: Text Mining

Large Scale Unstructured Document Classification Using Unlabeled Data and Syntactic Information .....	88
<i>Seong-Bae Park, Byoung-Tak Zhang</i>	

Extracting Shared Topics of Multiple Documents .....	100
<i>Xiang Ji, Hongyuan Zha</i>	

An Empirical Study on Dimensionality Optimization in Text Mining for Linguistic Knowledge Acquisition .....	111
<i>Yu-Seop Kim, Jeong-Ho Chang, Byoung-Tak Zhang</i>	

A Semi-supervised Algorithm for Pattern Discovery in Information Extraction from Textual Data .....	117
<i>Tianhao Wu, William M. Pottenger</i>	

## Session 2B: Bio Mining

Mining Patterns of Dyspepsia Symptoms Across Time Points Using Constraint Association Rules .....	124
<i>Annie Lau, Siew Siew Ong, Ashesh Mahidadia, Achim Hoffmann, Johanna Westbrook, Tatjana Zrimec</i>	

Predicting Protein Structural Class from Closed Protein Sequences .....	136
<i>N. Rattanakronkul, T. Wattarujeeekrit, K. Waiyamai</i>	

Learning Rules to Extract Protein Interactions from Biomedical Text ....	148
<i>Tu Minh Phuong, Doheon Lee, Kwang Hyung Lee</i>	

Predicting Protein Interactions in Human by Homologous Interactions in Yeast .....	159
<i>Hyongguen Kim, Jong Park, Kyungsook Han</i>	

## Session 2C: Web Mining

Mining the Customer's Up-To-Moment Preferences for E-commerce Recommendation .....	166
<i>Yi-Dong Shen, Qiang Yang, Zhong Zhang, Hongjun Lu</i>	

A Graph-Based Optimization Algorithm for Website Topology Using Interesting Association Rules .....	178
<i>Edmond H. Wu, Michael K. Ng</i>	

A Markovian Approach for Web User Profiling and Clustering .....	191
<i>Younes Hafri, Chabane Djeraba, Peter Stanchev, Bruno Bachimont</i>	

Extracting User Interests from Bookmarks on the Web .....	203
<i>Jason J. Jung, Geun-Sik Jo</i>	

## Session 3A: Stream Mining II

Mining Frequent Instances on Workflows .....	209
<i>Gianluigi Greco, Antonella Guzzo, Giuseppe Manco, Domenico Saccà</i>	

Real Time Video Data Mining for Surveillance Video Streams . . . . .	222
<i>JungHwan Oh, JeongKyu Lee, Sanjaykumar Kote</i>	
Distinguishing Causal and Acausal Temporal Relations . . . . .	234
<i>Kamran Karimi, Howard J. Hamilton</i>	

### Session 3B: Bayesian Networks

Online Bayes Point Machines . . . . .	241
<i>Edward Harrington, Ralf Herbrich, Jyrki Kivinen, John Platt, Robert C. Williamson</i>	
Exploiting Hierarchical Domain Values for Bayesian Learning . . . . .	253
<i>Yiqiu Han, Wai Lam</i>	
A New Restricted Bayesian Network Classifier . . . . .	265
<i>Hongbo Shi, Zhihai Wang, Geoff Webb, Houkuan Huang</i>	

### Session 3C: Clustering II

AGRID: An Efficient Algorithm for Clustering Large High-Dimensional Datasets . . . . .	271
<i>Zhao Yanchang, Song Junde</i>	
Multi-level Clustering and Reasoning about Its Clusters Using Region Connection Calculus . . . . .	283
<i>Ickjai Lee, Mary-Anne Williams</i>	
An Efficient Cell-Based Clustering Method for Handling Large, High-Dimensional Data . . . . .	295
<i>Jae-Woo Chang</i>	

### Session 4A: Association Rules I

Enhancing SWF for Incremental Association Mining by Itemset Maintenance . . . . .	301
<i>Chia-Hui Chang, Shi-Hsan Yang</i>	
Reducing Rule Covers with Deterministic Error Bounds . . . . .	313
<i>Vikram Pudi, Jayant R. Haritsa</i>	
Evolutionary Approach for Mining Association Rules on Dynamic Databases . . . . .	325
<i>P. Deepa Shenoy, K.G. Srinivasa, K.R. Venugopal, L.M. Patnaik</i>	

## Session 4B: Semi-structured Data Mining

Position Coded Pre-order Linked WAP-Tree for Web Log Sequential Pattern Mining .....	337
<i>Yi Lu, C.I. Ezeife</i>	
An Integrated System of Mining HTML Texts and Filtering Structured Documents .....	350
<i>Bo-Hyun Yun, Myung-Eun Lim, Soo-Hyun Park</i>	
A New Sequential Mining Approach to XML Document Similarity Computation .....	356
<i>Ho-pong Leung, Fu-lai Chung, Stephen Chi-fai Chan</i>	

## Session 4C: Classification I

Optimization of Fuzzy Rules for Classification Using Genetic Algorithm .....	363
<i>Myung Won Kim, Joung Woo Ryu, Samkeun Kim, Joong Geun Lee</i>	
Fast Pattern Selection for Support Vector Classifiers .....	376
<i>Hyunjung Shin, Sungzoon Cho</i>	
Averaged Boosting: A Noise-Robust Ensemble Method .....	388
<i>Yongdai Kim</i>	
Improving Performance of Decision Tree Algorithms with Multi-edited Nearest Neighbor Rule .....	394
<i>Chen-zhou Ye, Jie Yang, Li-xiu Yao, Nian-yi Chen</i>	

## Session 5A: Data Analysis

HOT: Hypergraph-Based Outlier Test for Categorical Data .....	399
<i>Li Wei, Weining Qian, Aoying Zhou, Wen Jin, Jeffrey X. Yu</i>	
A Method for Aggregating Partitions, Applications in K.D.D. ....	411
<i>Pierre-Emmanuel Jouve, Nicolas Nicoloyannis</i>	
Efficiently Computing Iceberg Cubes with Complex Constraints through Bounding .....	423
<i>Pauline L.H. Chou, Xiuzhen Zhang</i>	
Extraction of Tag Tree Patterns with Contractible Variables from Irregular Semistructured Data .....	430
<i>Tetsuhiro Miyahara, Yusuke Suzuki, Takayoshi Shoudai, Tomoyuki Uchida, Sachio Hirokawa, Kenichi Takahashi, Hiroaki Ueda</i>	

## Session 5B: Association Rules II

Step-By-Step Regression: A More Efficient Alternative for Polynomial Multiple Linear Regression in Stream Cube .....	437
<i>Chao Liu, Ming Zhang, Minrui Zheng, Yixin Chen</i>	
Progressive Weighted Miner: An Efficient Method for Time-Constraint Mining .....	449
<i>Chang-Hung Lee, Jian Chih Ou, Ming-Syan Chen</i>	
Mining Open Source Software (OSS) Data Using Association Rules Network .....	461
<i>Sanjay Chawla, Bavani Arunasalam, Joseph Davis</i>	
Parallel FP-Growth on PC Cluster .....	467
<i>Iko Pramudiono, Masaru Kitsuregawa</i>	

## Session 5C: Feature Selection

Active Feature Selection Using Classes .....	474
<i>Huan Liu, Lei Yu, Manoranjan Dash, Hiroshi Motoda</i>	
Electricity Based External Similarity of Categorical Attributes .....	486
<i>Christopher R. Palmer, Christos Faloutsos</i>	
Weighted Proportional k-Interval Discretization for Naive-Bayes Classifiers .....	501
<i>Ying Yang, Geoffrey I. Webb</i>	
Dealing with Relative Similarity in Clustering: An Indiscernibility Based Approach .....	513
<i>Shoji Hirano, Shusaku Tsumoto</i>	

## Session 6A: Stream Mining III

Considering Correlation between Variables to Improve Spatiotemporal Forecasting .....	519
<i>Zhigang Li, Liangang Liu, Margaret H. Dunham</i>	
Correlation Analysis of Spatial Time Series Datasets: A Filter-and-Refine Approach .....	532
<i>Pusheng Zhang, Yan Huang, Shashi Shekhar, Vipin Kumar</i>	
When to Update the Sequential Patterns of Stream Data? .....	545
<i>Qingguo Zheng, Ke Xu, Shilong Ma</i>	

## Session 6B: Clustering III

A New Clustering Algorithm for Transaction Data via Caucus .....	551
<i>Jinmei Xu, Hui Xiong, Sam Yuan Sung, Vipin Kumar</i>	

DBRS: A Density-Based Spatial Clustering Method  
with Random Sampling ..... 563  
*Xin Wang, Howard J. Hamilton*

Optimized Clustering for Anomaly Intrusion Detection ..... 576  
*Sang Hyun Oh, Won Suk Lee*

**Session 6C: Classification II**

Finding Frequent Subgraphs from Graph Structured Data with  
Geometric Information and Its Application to Lossless Compression ..... 582  
*Yuko Itokawa, Tomoyuki Uchida, Takayoshi Shoudai,  
Tetsuhiro Miyahara, Yasuaki Nakamura*

Upgrading ILP Rules to First-Order Bayesian Networks..... 595  
*Ratthachat Chatpatanasiri, Boonserm Kijirikul*

A Clustering Validity Assessment Index..... 602  
*Youngok Kim, Soowon Lee*

**Author Index** ..... 609

Advances in Knowledge Discovery and Data Mining  
7th Pacific-Asia Conference, PAKDD 2003. Seoul, Korea,  
April 30 - May 2, 2003, Proceedings  
Whang, K.-Y.; Jeon, J.; Shim, K.; Srivastava, J. (Eds.)  
2003, XVIII, 614 p., Softcover  
ISBN: 978-3-540-04760-5