

Tu Bao Ho David Cheung Huan Liu (Eds.)

Advances in Knowledge Discovery and Data Mining

9th Pacific-Asia Conference, PAKDD 2005
Hanoi, Vietnam, May 18-20, 2005
Proceedings

 Springer

Table of Contents

Keynote Speech and Invited Talks

Machine Learning for Analyzing Human Brain Function <i>Tom Mitchell</i>	1
Subgroup Discovery Techniques and Applications <i>Nada Lavrač</i>	2
IT Development in the 21 st Century and Its Implications <i>Unna Huh</i>	15

Theoretic Foundations

Data Mining of Gene Expression Microarray via Weighted Prefix Trees <i>Tran Trang, Nguyen Cam Chi, Hoang Ngoc Minh</i>	21
Automatic Extraction of Low Frequency Bilingual Word Pairs from Parallel Corpora with Various Languages <i>Hiroshi Echizen-ya, Kenji Araki, Yoshio Mornouchi</i>	32
A Kernel Function Method in Clustering <i>Ling Zhang, Tao Wu, Yanping Zhang</i>	38
Performance Measurements for Privacy Preserving Data Mining <i>Nan Zhang, Wei Zhao, Jianer Chen</i>	43
Extraction of Frequent Few-Overlapped Monotone DNF Formulas with Depth-First Pruning <i>Yoshikazu Shima, Kouichi Hirata, Masateru Harao</i>	50

Association Rules

Rule Extraction from Trained Support Vector Machines <i>Ying Zhang, HongYe Su, Tao Jia, Jian Chu</i>	61
Pruning Derivative Partial Rules During Impact Rule Discovery <i>Shiyong Huang, Geoffrey I. Webb</i>	71
IGB: A New Informative Generic Base of Association Rules <i>Gh. Gasmı, S. Ben Yahia, E. Mephu Nguifo, Y. Slimani</i>	81

A Divide and Conquer Approach for Deriving Partially Ordered Sub-structures <i>Sadok Ben Yahia, Yahya Slimani, Jihen Rezgui</i>	91
Finding Sporadic Rules Using Apriori-Inverse <i>Yun Sing Koh, Nathan Rountree</i>	97
Automatic View Selection: An Application to Image Mining <i>Manoranjan Dash, Deepak Kolippakkam</i>	107
Pushing Tougher Constraints in Frequent Pattern Mining <i>Francesco Bonchi, Claudio Lucchese</i>	114
An Efficient Compression Technique for Frequent Itemset Generation in Association Rule Mining <i>Mafruz Zaman Ashrafi, David Taniar, Kate Smith</i>	125
Mining Time-Profiled Associations: An Extended Abstract <i>Jin Soung Yoo, Pusheng Zhang, Shashi Shekhar</i>	136
Online Algorithms for Mining Inter-stream Associations from Large Sensor Networks <i>K. K. Loo, Ivy Tong, Ben Kao</i>	143
Mining Frequent Ordered Patterns <i>Zhi-Hong Deng, Cong-Rui Ji, Ming Zhang, and Shi-Wei Tang</i>	150
 Biomedical Domains	
Conditional Random Fields for Transmembrane Helix Prediction <i>Lior Lukov, Sanjay Chawla, W. Bret Church</i>	155
A DNA Index Structure Using Frequency and Position Information of Genetic Alphabet <i>Woo-Cheol Kim, Sanghyun Park, Jung-Im Won, Sang-Wook Kim, Jee-Hee Yoon</i>	162
An Automatic Unsupervised Querying Algorithm for Efficient Information Extraction in Biomedical Domain <i>Min Song, Il-Yeol Song, Xiaohua Hu, Robert B. Allen</i>	173
Voting Fuzzy k-NN to Predict Protein Subcellular Localization from Normalized Amino Acid Pair Compositions <i>Thai Quang Tung, Doheon Lee, Dae-Won Kim, Jong-Tae Lim</i>	180
Comparison of Tree Based Methods on Mammography Data <i>Richard De Veaux, Thu Hoàng</i>	186

Bayesian Sequence Learning for Predicting Protein Cleavage Points <i>Michael Mayo</i>	192
A Novel Indexing Method for Efficient Sequence Matching in Large DNA Database Environment <i>Jung-Im Won, Jee-Hee Yoon, Sanghyun Park, Sang-Wook Kim</i>	203
Classification and Ranking	
Threshold Tuning for Improved Classification Association Rule Mining <i>Frans Coenen, Paul Leng, Lu Zhang</i>	216
Using Rough Set in Feature Selection and Reduction in Face Recognition Problem <i>Le Hoai Bac, Nguyen Anh Tuan</i>	226
Analysis of Company Growth Data Using Genetic Algorithms on Binary Trees <i>Gerrit K. Janssens, Kenneth Sørensen, Arthur Limère, Koen Vanhoof</i>	234
Considering Re-occurring Features in Associative Classifiers <i>Rafal Rak, Wojciech Stach, Osmar R. Zaiane, Maria-Luiza Antonie</i>	240
A New Evolutionary Neural Network Classifier <i>Arit Thammano, Asavin Meengen</i>	249
A Privacy-Preserving Classification Mining Algorithm <i>Weiping Ge, Wei Wang, Xiaorong Li, Baile Shi</i>	256
Combining Classifiers with Multi-representation of Context in Word Sense Disambiguation <i>Cuong Anh Le, Va-Nam Huynh, Akira Shimazu</i>	262
Automatic Occupation Coding with Combination of Machine Learning and Hand-Crafted Rules <i>Kazuko Takahashi, Hiroya Takamura, Manabu Okumura</i>	269
Retrieval Based on Language Model with Relative Entropy and Feedback <i>Hua Huo, Boqin Feng</i>	280
Text Classification for DAG-Structured Categories <i>Cao D. Nguyen, Tran A. Dung, Tru H. Cao</i>	290
Sentiment Classification Using Word Sub-sequences and Dependency Sub-trees <i>Shotaro Matsumoto, Hiroya Takamura, Manabu Okumura</i>	301

Improving Rough Classifiers Using Concept Ontology <i>Nguyen Sinh Hoa, Nguyen Hung Son</i>	312
QED: An Efficient Framework for Temporal Region Query Processing <i>Yi-Hong Chu, Kun-Ta Chuang, Ming-Syan Chen</i>	323
Clustering	
A MPAA-Based Iterative Clustering Algorithm Augmented by Nearest Neighbors Search for Time-Series Data Streams <i>Jessica Lin, Michai Vlachos, Eamonn Keogh, Dimitrios Gunopulos, Jianwei Liu, Shoujian Yu, Jiajin Le</i>	333
Locating Motifs in Time-Series Data <i>Zheng Liu, Jeffrey Xu Yu, Xuemin Lin, Hongjun Lu, Wei Wang</i>	343
Stochastic Local Clustering for Massive Graphs <i>Satu Elisa Schaeffer</i>	354
A Neighborhood-Based Clustering Algorithm <i>Shuigeng Zhou, Yue Zhao, Jihong Guan, Joshua Huang</i>	361
Improved Self-splitting Competitive Learning Algorithm <i>Jun Liu, Kotagiri Ramamohanarao</i>	372
Speeding-Up Hierarchical Agglomerative Clustering in Presence of Expensive Metrics <i>Mirco Nanni</i>	378
Dynamic Cluster Formation Using Level Set Methods <i>Andy M. Yip, Chris Ding, Tony F. Chan</i>	388
A Vector Field Visualization Technique for Self-organizing Maps <i>Georg Pözlbauer, Andreas Rauber, Michael Dittenbach</i>	399
Visualization of Cluster Changes by Comparing Self-organizing Maps <i>Denny McG. Scurie, David McG. Squire</i>	410
An Incremental Data Stream Clustering Algorithm Based on Dense Units Detection <i>Jing Gao, Jianzhong Li, Zhaogong Zhang, Pang-Ning Tan</i>	420
Visual Interactive Evolutionary Algorithm for High Dimensional Data Clustering and Outlier Detection <i>Lydia Boudjeloud, François Poulet</i>	426

Approximated Clustering of Distributed High-Dimensional Data <i>Hans-Peter Kriegel, Peter Kunath, Martin Pfeifle, Matthias Renz</i>	432
--	-----

Dynamic Data Mining

Improvements of IncSpan: Incremental Mining of Sequential Patterns in Large Database <i>Son N. Nguyen, Xingzhi Sun, Maria E. Orlowska</i>	442
Efficient Sampling: Application to Image Data <i>Surong Wang, Manoranjan Dash, Liang-Tien Chia</i>	452
Cluster-Based Rough Set Construction <i>Qiang Li, Bo Zhang</i>	464

Graphic Model Discovery

Learning Bayesian Networks Structures from Incomplete Data: An Efficient Approach Based on Extended Evolutionary Programming <i>Xiaolin Li, Xiangdong He, Senmiao Yuan</i>	474
Dynamic Fuzzy Clustering for Recommender Systems <i>Sung-Hwan Min, Ingoo Han</i>	480
Improving Mining Quality by Exploiting Data Dependency <i>Fang Chu, Yizhou Wang, Carlo Zaniolo, D. Stott Parker</i>	486

High Dimensional Data

Feature Selection for High Dimensional Face Image Using Self-organizing Maps <i>Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, Fuyan Zhang</i>	500
Progressive Sampling for Association Rules Based on Sampling Error Estimation <i>Kun-Ta Chuang, Ming-Syan Chen, Wen-Chieh Yang</i>	505
<i>CLeVer</i> : A Feature Subset Selection Technique for Multivariate Time Series <i>Kiyoung Yang, Hyunjin Yoon, Cyrus Shahabi</i>	516
Covariance and PCA for Categorical Variables <i>Hiroataka Niitsuma, Takashi Okada</i>	523

Integration of Data Warehousing

ADenTS: An Adaptive Density-Based Tree Structure for Approximating Aggregate Queries over Real Attributes <i>Tianyì Wu, Jian Xu, Chen Wang, Wei Wang, Baile Shi</i>	529
--	-----

Frequent Itemset Mining with Parallel RDBMS <i>Xuequn Shang, Kai-Uwe Sattler</i>	539
---	-----

Knowledge Management

Using Consensus Susceptibility and Consistency Measures for Inconsistent Knowledge Management <i>Ngoc Thanh Nguyen, Michal Malowicki</i>	545
---	-----

WLPMiner: Weighted Frequent Pattern Mining with Length-Decreasing Support Constraints <i>Unil Yun, John J. Leggett</i>	555
---	-----

Machine Learning Methods

A Framework for Incorporating Class Priors into Discriminative Classification <i>Rong Jin, Yi Liu</i>	568
--	-----

Increasing Classification Accuracy by Combining Adaptive Sampling and Convex Pseudo-Data <i>Chia Huey Ooi, Madhu Chetty</i>	578
--	-----

Kernels over Relational Algebra Structures <i>Adam Woźnica, Alexandros Kalousis, Melanie Hilario</i>	588
---	-----

Adaptive Nonlinear Auto-Associative Modeling Through Manifold Learning <i>Junping Zhang, Stan Z. Li</i>	599
--	-----

Maximizing Tree Diversity by Building Complete-Random Decision Trees <i>Fei Tony Liu, Kai Ming Ting, Wei Fan</i>	605
---	-----

SETRED: Self-training with Editing <i>Ming Li, Zhi-Hua Zhou</i>	611
--	-----

Adjusting Mixture Weights of Gaussian Mixture Model via Regularized Probabilistic Latent Semantic Analysis <i>Luo Si, Rong Jin</i>	622
---	-----

Training Support Vector Machines Using Greedy Stagewise Algorithm <i>Liefeng Bo, Ling Wang, Licheng Jiao</i>	632
CI-GBI: A Novel Approach for Extracting Typical Patterns from Graph-Structured Data <i>Phu Chien Nguyen, Kouzou Ohara, Hiroshi Motoda, Takashi Washio</i>	639
Improved Bayesian Spam Filtering Based on Co-weighted Multi-area Information <i>Raju Shrestha, Yaping Lin</i>	650
Novel Algorithms	
An Efficient Framework for Mining Flexible Constraints <i>Arnaud Soulet, Bruno Crémilleux</i>	661
Support Oriented Discovery of Generalized Disjunction-Free Representation of Frequent Patterns with Negation <i>Marzena Kryszkiewicz, Katarzyna Cichoń</i>	672
Feature Selection Algorithm for Data with Both Nominal and Continuous Features <i>Wenyin Tang, Kezhi Mao</i>	683
A Two-Phase Algorithm for Fast Discovery of High Utility Itemsets <i>Ying Liu, Wei-keng Liao, Alok Choudhary</i>	689
On Multiple Query Optimization in Data Mining <i>Marek Wojciechowski, Maciej Zakrzewicz</i>	696
USAID: Unifying Signature-Based and Anomaly-Based Intrusion Detection <i>Zhuowei Li, Amitabha Das, Jianying Zhou</i>	702
Spatial Data	
Mining Mobile Group Patterns: A Trajectory-Based Approach <i>San-Yih Hwang, Ying-Han Liu, Jeng-Kuen Chiu, Ee-Peng Lim</i>	713
Can We Apply Projection Based Frequent Pattern Mining Paradigm to Spatial Co-location Mining? <i>Yan Huang, Liqin Zhang, Ping Yu</i>	719
PatZip: Pattern-Preserved Spatial Data Compression <i>Yu Qian, Kang Zhang, D. T. Huynh</i>	726

Dynamic Mining Hierarchical Topic from Web News Stream Data Using Divisive-Agglomerative Clustering Method <i>Jian-Wei Liu, Shou-Jian Yu, Jia-Jin Le</i>	826
Collecting Topic-Related Web Pages for Link Structure Analysis by Using a Potential Hub and Authority First Approach <i>Leuo-Hong Wang, Tong-Wen Lee</i>	832
A Top-Down Algorithm for Mining Web Access Patterns from Web Logs <i>Jian-Kui Guo, Bei-jun Ruan, Zun-ping Cheng, Fang-zhong Su, Ya-qin Wang, Xu-bin Deng, Shang Ning, Yang-Yong Zhu</i>	838
Kernel Principal Component Analysis for Content Based Image Retrieval <i>Guang-Ho Cha</i>	844
Mining Frequent Trees with Node-Inclusion Constraints <i>Atsuyoshi Nakamura, Mineichi Kudo</i>	850
Author Index	861