Tentative Program of PAKDD 2005

Wednesday, May 18th 2005

8:30-18:30	*Registration*	
9:00-12:00	*Workshop A* *Room:* Ballroom 1	**Knowledge Discovery and Data Management in Biomedical Science**
9:00-12:00	*Tutorial A* *Room:* Function Room 3	**Graph Mining Techniques and Their Applications**
	Tutorial B *Room:* Function Room 7	**Rough Set Approach to KDD**
12:00-13:30	Lunch	
13:30-16:30	*Workshop A* *Room:* Ballroom 1	**Knowledge Discovery and Data Management in Biomedical Science**
	Workshop B *Room:* Ballroom 2	**Rough Set Techniques in Knowledge Discovery**
13:30-16:30	*Tutorial C* *Room:* Function Room 7	**Advanced Techniques for Information and Image Classification for Knowledge Management and Decision Making**
18:00-20:00	*Reception*	

Thursday, May 19th 2005

8:00-18:00	Registration
8:30-9:00	Opening
9:00-10:00	*Keynote speech* *Chair: Tu Bao Ho, Room:* Ballroom 1 **Machine Learning for Analyzing Human Brain Function** *Tom Mitchell*
10:00-10:25	Coffee Break
10:25-12:00	*Session 1A: Novel Algorithms* *Chair: Takashi Washio, Room:* Ballroom 1 **An Efficient Framework for Mining Flexible Constraints** (R) **BEST STUDENT PAPER** *Arnaud Soulet, Bruno Crémilleux* **Support Oriented Discovery of Generalized Disjunction-Free Representation of Frequent Patterns with Negation** (R) *Marzena Kryszkiewicz, Katarzyna Cichon* **Feature Selection Algorithm for Data with Both Nominal and Continuous Features** *Wenyin Tang, Kezhi Mao* (S) **A Two-Phase Algorithm for Fast Discovery of High Utility Itemsets** (S) *Ying Liu, Wei-keng Liao, Alok Choudhary* **On Multiple Query Optimization in Data Mining** (S) *Marek Wojciechowski, Maciej Zakrzewicz* *Session 1B: Biomedical Domains* *Chair: Kenji Satou, Room:* Ballroom 2 **Bayesian Sequence Learning For Predicting Protein Cleavage Points** (R) *Michael Mayo* **A Novel Indexing Method for Efficient Sequence Matching in Large DNA Database Environment** (R) *Jung-Im Won, Jee-Hee Yoon, Sanghyun Park, Sang-Wook Kim* **An Automatic Unsupervised Querying Algorithm for Efficient Information Extraction in Biomedical Domain** (S) *Min Song, Il-Yeol Song, Xiaohua Hu, Robert Allen* **Voting Fuzzy K-NN to Predict Protein Subcellular Localization from Normalized Amino Acid Pair Compositions** (S) *Thai Quang Tung, Doheon Lee, Dae-Won Kim, Jong-Tae Lim* **Comparison of Tree based methods on Mammography Data** (S) *Richard De Veaux, Thu Hoang*

	Session 1C: Text and Web Data Mining *Chair:* *Ee-Peng Lim*, *Room:* Function Room 7 **Subspace Clustering of Text Documents with Feature Weighting K-Means Algorithm** *Liping Jing, Michael K. Ng, Jun Xu, Joshua Zhexue Huang* (R) **Mining Frequent Trees with Node-Inclusion Constraints** (R) *Atsuyoshi Nakamura, Mineichi Kudo* **Using Term Clustering and Supervised Term Affinity Construction o Boost Text Classification** (S) *Chong Wang, Wenyuan Wang* **Technology Trends Analysis from the Internet Resources** (S) *Shin-ichi Kobayashi, Yasuyuki Shirai, Kazuo Hiyane, Fumihiro Kumeno, Hiroshi Inujima, Noriyoshi Yamauchi* **Dynamic Mining Hierarchical Topic from Web News Stream Data using Divisive-Agglomerative Clustering Method** (S) *Jian-Wei Liu, Shou-Jian Yu, Jia-Jin Le* *Session 1D: Machine Learning Methods* *Chair:* *Frans Coenen*, *Room:* Function Room 3 **A Framework for Incorporating Class Priors into Discriminative Classification** (R) *Rong Jin, Yi Liu* **Improved Bayesian Spam Filtering Based on Co-weighted Multi-area Information** (R) *Raju Shrestha, Yaping Lin* **Adaptive Nonlinear Auto-Associative Modeling through Manifold Learning with Applications for Character and Digit Recognition** (S) *Junping Zhang, Stan Z. Li* **Maximizing Tree Diversity by Building Complete-Random Decision Trees** (S) *Fei Tony Liu, Kai Ming Ting, Wei Fan* **Training Support Vector Machines Using Greedy Stagewise Algorithm** (S) *Liefeng Bo, Ling Wang, Licheng Jiao*
12:00-13:30	Lunch
13:30-14:20	*Invited talk* *Chair:* *David Cheung*, *Room:* Ballroom 1 **IT development in the 21st Century and Its Implications** Unna Huh

14:20-15:00	*Sessions 2A: Integration of Data Warehousing* *Chair:* Marcin Szczuka, *Room:* Function Room 3 **ADenTS: An Adaptive Density-based Tree Structure for Approximating Aggregate Queries over Real Attributes** (R) *Tianyi Wu, Jian Xu, Chen Wang, Wei Wang, Baile Shi* **Frequent Itemset Mining with Parallel RDBMS** (S) *Xuequn Shang, Kai-Uwe Sattler* *Session 2B: Biomedical Domains* *Chair:* Kouzou Ohara, *Room:* Ballroom 2 **A DNA Index Structure Using Frequency and Position Information of Genetic Alphabet** (R) *Woo-Cheol Kim, Sanghyun Park, Jung-Im Won, Sang-Wook Kim, Jee-Hee Yoon* **Conditional Random Fields for Transmembrane Helix Prediction** (S) *Lior Lukov, Sanjay Chawla, W. Bret Church* *Session 2C: Temporal Data* *Chair:* Gerrit K. Janssens, *Room:* Function Room 7 **A Likelihood Ratio Distance Measure for the Similarity between the Fourier Transform of Time Series** (S) *Anthony Bagnall, Gareth Janacek, Michael Powell* **The TIMERS II Algorithm for the Discovery of Causality** (S) *Howard J. Hamilton, Kamran Karimi* **A Recent-Based Dimension Reduction Technique for Time Series Data** (S) *Yanchang Zhao, Chengqi Zhang, Shichao Zhang* *Session 2D: Text and Web Data Mining* *Chair:* Rao Kotagiri, *Room:* Ballroom 1 **Collecting Topic-related Web Pages for Link Structure Analysis by Using a Potential Hub and Authority First Approach** (S) *Leuo-hong Wang, Tong-wen Lee* **A Top-down Algorithm for Mining Web Access Patterns from Web Logs** (S) *Guo Jian-Kui, Ruan Bei-jun, Cheng Zun-ping, Su Fang-zhong, Wang Ya-qin, Deng Xu-bin, Shang Ning, Zhu Yang-Yong* **Kernel Principal Component Analysis for Content Based Image Retrieval** (S) *Guang-Ho Cha*

15:00-15:30	Coffee Break
15:30-17:15	*Session 3A: Theoretic Foundations* *Chair:* *Graham Williams*, *Room:* Ballroom 1 **Data Mining of Gene Expression Microarray via Weighted Prefix Trees** (R) *Tran Trang, Nguyen Cam Chi, Hoang Ngoc Minh* **A Kennel Function Method in Clustering** (R) *Ling Zhang, Tao Wu, Yanping Zhang* **Extraction of Frequent Few-Overlapped Monotone DNF Formulas with Depth-First Pruning** (R) *Yoshikazu Shima, Kouichi Hirata, Masateru Harao* **Automatic Extraction of Low Frequency Bilingual Word Paris from Parallel Corpora with Various Languages** (S) *Hiroshi Echizen-ya, Kenji Araki, Yoshio Mornouchi* **Performance Measurements for Privacy Preserving Data Mining** (S) *Nan Zhang,Wei Zhao, Jianer Chen* *Session 3B: Classification and Ranking* *Chair:* *Ning Zhong*, *Room:* Ballroom 2 **Threshold Tuning for Improved Classification Association Rule Mining** (R) *Frans Coenen, Paul Leng, Lu Zhang* **Automatic Occupation Coding with Combination of Machine Learning and Hand-Crafted Rules** (R) *Kazuko Takahashi, Hiroya Takamura, Manabu Okumura* **Retrieval Based on Language Model with Relative Entropy and Feedback** (R) *Hua Huo, Boqin Feng* **Using Rough Set in Feature Selection and Reduction in Face Recognition Problem** *Le Hoai Bac, Nguyen Tuan Anh* (S) **Analysis of company growth data using genetic algorithms on binary trees** (S) *Gerrit K. Janssens, Kenneth Sörensen, Arthur Limère, Koen Vanhoof*

	Session 3C: Clustering *Chair:* *Joshua Z. Huang*, *Room:* Function Room 7 **A MPAA-Based Iterative Clustering Algorithm augmented by Nearest Neighbors Search for Time-Series Data Streams** (R) *Jessica Lin, Michai Vlachos, Eamonn Keogh, Dimitrios Gunopulos, Jian-Wei Liu, Shou-Jian Yu, Jia-Jin Le* **A Neighborhood-Based Clustering Algorithm** (R) *Shuigeng Zhou, Yue Zhao, Jihong Guan, Joshua Huang* **Locating Motifs in Time-Series Data** (R) *Zheng Liu, Jeffrey Xu Yu, Xuemin Lin, Hongjun Lu, Wei Wang* **Stochastic local clustering for massive graphs** (S) *Satu Elisa Schaeffer* **Improved Self-Splitting Competitive Learning Algorithm** (S) *Jun Liu, Kotagiri Ramamohanarao* *Session 3D: Association Rules* *Chair:* *Hoang Tru Cao*, *Room:* Function Room 3 **Rule Extraction from Trained Support Vector Machines** (R) *Ying Zhang, HongYe Su, Tao Jia, Jian Chu* **Pruning Derivative Partial Rules during Impact Rule Discovery** (R) *Shiying Huang, Geoffrey I. Webb* **IGB: A New Informative Generic Base of Association Rules** (R) *Gh. Gasmi, S. Ben Yahia, E. Mephu Nguifo, Y. Slimani* **A Divide and Conquer Approach for Deriving Partially Ordered Sub-structures** (S) *Sadok Ben Yahia, Yahya Slimani, Jihen Rezgui* **Automatic View Selection: An Application to Image Mining** (S) *Manoranjan Dash, Deepak Kolippakkam*
18:30-21:00	*Banquet*

Friday, May 20th 05

8:00-12:00	Registration
9:00-09:50	*Invited talk* *Chair:* *Hiroshi Motoda*, *Room:* Ballroom 1 **Subgroup Discovery: Techniques and Applications** *Nada Lavrac*
9:50-10:20	Coffee Break
10:20-12:00	*Session 4A: Machine Learning Methods* *Chair:* *San-Yih Hwang*, *Room:* Ballroom 1 **Kernels over relational algebra structures** (R) **BEST STUDENT PAPER** *Adam Woznica, Alexandros Kalousis, Melanie Hilario* **SETRED: Self-Training with Editing** (R) *Ming Li, Zhi-Hua Zhou* **CI-GBI: A Novel Approach for Extracting Typical Patterns from Graph-Structured Data** (R) *Phu Chien Nguyen, Kouzou Ohara, Hiroshi Motoda, Takashi Washio* **Adjusting Mixture Weights of Gaussian Mixture Model via Regularized Probabilistic Latent Semantic Analysis** (R) *Luo Si, Rong Jin* *Session 4B: Association Rules* *Chair:* *Osmar Zaiane*, *Room:* Ballroom 2 **Finding Sporadic Rules Using Apriori-Inverse** (R) *Yun Sing Koh, Nathan Rountree* **Pushing Tougher Constraints in Frequent Pattern Mining** (R) *Francesco Bonchi, Claudio Lucchese* **Mining Time-Profiled Associations: An Extended Abstract** (S) *Jin Soung Yoo, Pusheng Zhang, Shashi Shekhar* **Online Algorithms for Mining Inter-Stream Associations from Large Sensor Neworks** *K. K. Loo, Ivy Tong, Ben Kao* (S) **Mining Frequent Ordered Patterns** (S) *Zhi-Hong Deng, Cong-Rui Ji, Ming Zhang, and Shi-Wei Tang*

	Session 4C: Classification and Ranking *Chair:* Martin Pfeifle, *Room:* Function Room 7 **Text Classification for DAG-Structured Categories** (R) *Cao D. Nguyen, Tran A. Dung, Tru H. Cao* **Sentiment Classification using Word Sub-Sequences and Dependency Sub-Trees** (R) *Shaotaro Matsumoto, Hiroya Takamura, Manabu Okumura* **A New Evolutionary Neural Network** Classifier (S) *Arit Thammano, Asavin Meengen* **Combining Classifiers with Multi-Representation of Context in Word Sense Disambiguation** (S) *Cuong Anh Le, Van Nam Huynh, Akira Shimazu* **A Privacy-Preserving Classification Mining Algorithm** (S) *Weiping Ge, Wei Wang, Xiaorong Li, Baile Shi* *Session 4D: High Dimensional Data* *Chair:* Tamas Horvath, *Room:* Function Room 3 **Progressive Sampling for Association Rules based on Sampling Error Estimation** (R) *Kun-Ta Chuang, Ming-Syan Chen, Wen-Chieh Yang* **CLeVer: A Feature Subset Selection Technique for Multivariate Time Series** (S) *Kiyoung Yang, Hyunjin Yoon, Cyrus Shahabi* **Covariance and PCA for Categorical Variables** (S) *Hirotaka Niitsuma, Takashi Okada* **Feature Selection for High Dimensional Face Image Using Self-Organizing Maps** (S) *Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, Fuyan Zhang*
12:00-13:30	Lunch
13:30-15:05	*Session 5A: Clustering* *Chair:* Zhi-Hua Zhou, *Room:* Function Room 3 **Speeding-up Hierarchical Agglomerative Clustering in Presence of Expensive Metrics** (R) *Mirco Nanni* **Dynamic Cluster Formation using Level Set Methods** (R) *Andy M. Yip, Chris Ding, Tony F. Chan* **An Incremental Data Stream Clustering Algorithm Based on Dense Units Detection** *Jing Gao, Jianzhong Li, Zhaogong Zhang, Pang-Ning Tan* (S)

Visual Interactive Evolutionary Algorithm for High Dimensional Data Clustering and Outlier Detection (S)

Lydia Boudjeloud, François Poulet

Session 5B: Spatial Data & Association Rules

Chair: Howard Hamilton, Room: Function Room 4

PatZip: Pattern-Preserved Spatial Data Compression (R)

Yu Qian, Kang Zhang, D. T. Huynh

An Efficient Compression Technique for Frequent Itemset Generation in Association Rule Mining (R)

Mafruz Zaman Ashrafi, David Taniar, Kate Smith

Mining Mobile Group Patterns: A Trajectory-based Approach (S)

San-Yih Hwang, Ying-Han Liu, Jeng-Kuen Chiu, Ee-Peng Lim

Can We Apply Projection Based Frequent Pattern Mining Paradigm to Spatial Co-location Mining? (S)

Yan Huang, Liqin Zhang, Ping Yu

Session 5C: Classification and Ranking

Chair: Arit Thammano, Room: Function Room 6

Improving Rough Classifiers Using Concept Ontology (R)

Sinh Hoa Nguyen, Hung Son Nguyen

ED: An Efficient Framework for Temporal Region Query Processing (R)

Yi-Hong Chu, Kun-Ta Chuang, Ming-Syan Chen

Increasing Classification Accuracy by Combining Adaptive Sampling and Convex Pseudo-Data (R)

Chia Huey Ooi, Madhu Chetty

Considering Re-occurring Features in Associative Classifiers (S)

Rafal Rak, Wojciech Stach, Osmar R. Za¨ıane, Maria-Luiza Antonie

	Session 5D: Knowledge Management & Novel Algorithms *Chair:* *Marzena Kryszkiewicz*, *Room:* Function Room 7 **Using Consensus Susceptibility and Consistency Measures for Inconsistent Knowledge Management** (R) *Ngoc Thanh Nguyen, Michal Malowiecki* **WLPMiner: Weighted Frequent Pattern Mining with Length-decreasing Support Constraints** (R) *Unil Yun, John J. Leggett* **USAID: Unifying Signature-based and Anomaly-based Intrusion Detection** (R) *Zhuowei Li, Amitabha Das, Jianying Zhou*
15:05-15:30	Coffee Break
15:30-17:00	**Session 6A:** *Temporal Data* *Chair:* *Takehisa Yairi*, *Room:* Function Room 3 **Cyclic Pattern Kernels Revisited** (R) *Tamas Horvath* **Accurate Symbolization of Time Series** (S) *Xinqiang Zuo, Xiaoming Jin* **A Novel Bit Level Time Series Representation with Implication of Similarity Search and Clustering** (S) *Chotirat Ratanamahatana, Eamonn Keogh, Anthony J. Bagnall, Stefano Lonardi* **Finding temporal features of Event-oriented patterns** (S) *Xingzhi Sun, Maria E. Orlowska, Xue Li* **An Anomaly Detection Method for Spacecraft using Relevance Vector Learning** (S) *Ryohei Fujimaki, Takehisa Yairi, Kazuo Machida* *Session 6B: Dynamic Data Mining* *Chair:* *Nguyen Hung Son*, *Room:* Function Room 4 **Improvements of IncSpan: Incremental Mining of Sequential Patterns in Large Database** (R) *Son N. Nguyen, Xingzhi Sun, Maria E. Orlowska* **Efficient Sampling: Application to Image Data** (R) *Surong Wang, Manoranjan Dash, Liang-Tien Chia* **Cluster-based Rough Set Construction** (R) *Qiang Li, Bo Zhang*

Session 6C: Graphic Model Discovery

Chair: *Takashi Okada*, *Room:* Function Room 6

Improving Mining Quality by Exploiting Data Dependency (R)

Fang Chu, Yizhou Wang, Carlo Zaniolo, D.Stott Parker

Learning Bayesian Networks Structures from Incomplete Data: An Efficient Approach Based on Extended Evolutionary Programming (S)

Xiaolin Li, Xiangdong He, Senmiao Yuan

Dynamic Fuzzy Clustering for Recommender Systems (S)

Sung-Hwan Min, Ingoo Han

Graph Partition Model for Robust Temporal Data Segmentation (S)

Yuan Jinhui, Zhang Bo, Lin Fuzong

Session 6D: Clustering

Chair: *Chengqi Zhang*, *Room:* Function Room 7

A vector field visualization technique for Self-Organizing Maps (R)

Georg Pölzlbauer, Andreas Rauber, Michael Dittenbach

Visualization of Cluster Changes by Comparing Self-Organizing Maps (R)

Denny, David McG. Squire

Approximated Clustering of Distributed High-Dimensional Data (R)

Peter Kunath