

Conference at a Glance

Sunday 27 November 2005

Tutorials			
Coffee breaks:10:00 - 10:30 and 15:00 - 15:30. Lunch break: 12:00 - 13:30			
ID	Tutorial Title	Room	Time
T1	Invited Tutorial: Models and Methods for Privacy-Preserving Data Mining and Data Publishing	Champions VI, VII	9:00 am - 12:00 pm
T2	Clustering with Constraints	Champions I, II	9:00 am-12:00 pm
T3	Bioinformatics and bioimage analysis	Champions VI, VII	1:30 pm - 4:50 pm
T4	Invited Tutorial: Mining and searching of graph-structured databases	Champions I, II	1:30 pm - 4:30 pm
T5	Invited Tutorial: DMX, XML for Analysis and SQL Server Data Mining Platform Note: this tutorial is complimentary for all conference participants (without any charges)	Legends I, II, III, IV	7:00 pm - 10:00 pm

Workshops			
Coffee breaks:10:00-10:30 and 15:00-15:30. Lunch break: 12:45-13:15			
ID	Workshop Title	Room	Time
W1	Mining Complex Data	Champions III	8:00 am-12:45 pm
W2	Data Mining Case Studies and ICDM Data Mining Practice Prizes (sponsored by Elder Research Inc.)	Founders Ballroom 2	8:00 am-12:45 pm
W3	Optimization-based Data Mining Techniques with Applications	Founders Ballroom 3	8:00 am-12:45 pm
W4	Multiagent Data Warehousing and Multiagent Data Mining	Champions V	8:00 am-12:45 pm
W5	Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources	Founders Ballroom 3	1:15 pm - 6:00 pm
W6	Privacy and Security Aspects of Data Mining	Champions V	1:15 pm - 6:00 pm
W7	Computational Intelligence in Data Mining	Champions III	1:15 pm - 6:00 pm
W8	Foundation of Semantic Oriented Data and Web Mining	Founders Ballroom 4	8:00 am - 6:00 pm
W9	Temporal data mining: algorithms, theory and applications	Founders Ballroom 1	8:00 am - 6:00 pm

Monday 28 November 2005

8:30 am - 9:00 am	Open Session Legends I, II, III, IV		
9:00 am – 10:00 am	Keynote Speech Dr. Raj Reddy (Turing Award Winner): "The Million Book Digital Library Project: Research Problems in Data Mining And Discovery" more info. Legends I, II, III, IV		
10:00 am – 10:30 am	Coffee Break Legends IV and Prefunction		
10:30 am - 1:00 pm	Paper Session 1: Times Series I Champions I, II	Paper Session 2: Clustering Schemes I Discovery Center A	Paper Session 3: Pattern Analysis on Text and Structured Data Discovery Center B
1:00 pm - 2:00 pm	Lunch Break Lunch provided by the conference		
2:00 pm - 3:00 pm	Keynote Speech Dr. John F. Elder IV: "Top 10 Data Mining Mistakes" more info. Legends I, II, III, IV		
3:00 pm - 3:30 pm	Coffee Break Legends IV and Prefunction		
3:30 pm - 6:00 pm	Paper Session 4: Times Series II Champions I,II	Paper Session 5: Clustering Schemes II Discovery Center A	Paper Session 6: Quality Assessment Discovery Center B

Tuesday 29 November 2005

9:00 am – 10:00 am	Keynote Speech Dr. Sunita Sarawagi: "Graphical Models for Structure Extraction and Information Integration" more info. Legends I, II, III, IV		
10:00 am – 10:15 am	Coffee Break Legends IV and Prefunction		
10:15 am – 12:45 pm	Paper Session 7: Times Series III Champions I, II	Paper Session 8: Spatial Data and Classification Schemes Discovery Center A	Paper Session 9: Preprocessing Techniques and Feature Selection Discovery Center B
12:45 pm – 2:15 pm	Lunch Break: (Buy your own lunch!)		
2:15 pm – 3:15 pm	Panel Session Legends I, II, III, IV		
3:15 pm – 3:30 pm	Coffee Break Legends IV and Prefunction		

3:30 pm – 6:00 pm	Paper Session 10: Learning Techniques I Champions I,II	Paper Session 11: Data Representation Discovery Center A	Paper Session 12: Security and Privacy Discovery Center B
7:15 pm	Banquet at NASA Houston Space Center		

Wednesday 30 November 2005

9:00 am - 10:00 am	Keynote Speech Dr. Arie Shoshani: "Efficient Indexing Technology for Data Mining of Scientific Data" more info. Legends I, II, III, IV		
10:00 am - 10:30 am	Coffee Break Legends IV and Prefunction		
10:30 am - 1:00 pm	Paper Session 13: Learning Techniques II Champions I,II	Paper Session 14: Data Mining Applications: Bio-Medical and Social Discovery Center A	Paper Session 15: Statistical Methods I Discovery Center B
1:00 pm - 2:00 pm	Lunch Break Box lunch provided by the conference	ICDM Business Meeting (1:15 pm –2:00 pm) Legends I, II, III, IV	
2:00 pm - 3:30 pm	Paper Session 16: Learning Techniques III Champions I,II	Paper Session 17: Data Mining Applications: Web Discovery Center A	Paper Session 18: Statistical Methods II Discovery Center B
3:00 pm - 4:00 pm	Coffee Break Legends IV and Prefunction		
4:00 pm - 5:30 pm	Paper Session 19: Tools and Algorithms Champions I, II	Paper Session 20: Data Mining Applications Discovery Center A	Paper Session 21: Optimization Techniques Discovery Center B

ICDM'05 Sunday November 27, 2005

Tutorials

There are coffee breaks from 10:00-10:30 and from 15:00-15:30. The lunch break is from 12:00-13:30 (Lunch is not included in the registration fee).

Morning Tutorials (9:00 am - 12:00 pm)

- Champions VI, VII: Models and Methods for Privacy-Preserving Data Mining and Data Publishing, *Johannes Gehrke* (Cornell University)

- Champions I,II: Clustering with Constraints, *Sugato Basu* (SRI International) and *Ian Davidson* (University At Albany, State University of New York)

Afternoon Tutorials(1:30 pm - 4:30 pm)

- Champions VI,VII: Bioinformatics and bioimage analysis, *Chris Ding and Hanchuan Peng* (* end at 4:50 pm)
- Champions I,II: Mining and searching of graph-structured databases, *Jiawei Han* , *Xifeng Yan* (University of Illinois at Urbana-Champaign), and *Philip Yu* (IBM Thomas J. Watson Research Center)

Evening Tutorials (7:00 pm - 10:00 pm)

- Legends I, II, III, IV: DMX, XML for Analysis and SQL Server Data Mining Platform, *Zhaohui Tang* (Microsoft Corp.)

Workshops

All Day Workshops (8:00 am - 6:00 pm)

There are coffee breaks from 10:00-10:30 and from 15:00-15:30. The lunch break is from 12:45-13:15 (Lunch is not included in the registration fee).

- Founders Ballroom 4: Foundation of Semantic Oriented Data and Web Mining, Organizers: *T. Y. Lin, S. Smale, Anita Wasilewska, Tomaso Poggio, Fred Petry and Ying Xie*
- Founders Ballroom 1: Temporal data mining: algorithms, theory and applications, Organizers: *Sheng Ma, Tao Li and Charles Perng*

Morning Workshops (8:00 am - 12:45 pm)

- Champions III: Mining Complex Data, Organizers: *Djamel A. Zighed, Shusaku Tsumoto and Zbigniew W. Ras*
- Founders Ballroom 2: Data Mining Case Studies and ICDM Data Mining Practice Prizes, Organizers: *Brendan Kitts, Gabor Melli*
- Founders Ballroom 3: Optimization-based Data Mining Techniques with Applications, Organizers: *Yong Shi*
- Champions V: Multiagent Data Warehousing and Multiagent Data Mining, Organizers: *M. N. Huhns, Wen-Ran Zhang, Yan-Qing Zhang and Xiaohua Tony Hu*

Afternoon Workshops (1:15 pm - 6:00 pm)

- Founders Ballroom 3: Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, Organizers: *Doina Caragea, Vasant Honavar, Ion Muslea and Raghu Ramakrishnan*
- Champions V: Privacy and Security Aspects of Data Mining, Organizers: *Stan Matwin and LiWu Chang*

- *Champions III: Computational Intelligence in Data Mining, Organizers: Fernando Berzal, Juan-Carlos Cubero, Zbigniew W. Ras, Thomas Sudkamp and Ronald R. Yager*

ICDM 2005 Monday, November 28, 2005

8:30 am - 9:00 am Open Session (Legends I,II,III,IV)

9:00 am - 10:00 am Keynote Speech (Legends I,II,III,IV)

The Million Book Digital Library Project: Research Problems in Data Mining And Discovery
Dr. Raj Reddy (Carnegie Mellon University, USA, Turing Award Winner)

Creating a universal, free to read, digital library containing all the books ever published is technically feasible today. Google, Yahoo and Microsoft have all announced their intention to scan and make available books of interest to public. Unfortunately many of these will be in English and inaccessible to over 80% of the world's population. Even when books in other languages become available online, their content will remain incomprehensible to most people. Natural Language Processing Technology is not yet perfect but promises to provide a way out of this conundrum. In this talk, we will discuss some of the special and unique research problems in data discovery arising in digital libraries and other online content, such as multi-lingual search, translation and summarization.

10:00 am - 10:30 am Coffee Break

10:30 am - 1:00 pm Paper Sessions (3 parallel Tracks)

Paper Session 1: Times Series I (Champions I, II)

(4 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Charles X. Ling (dr_charles_ling@yahoo.com)

- (R) Modeling Multiple Time Series for Anomaly Detection, by Philip Chan and Matthew Mahoney
- (R) Integrating Hidden Markov Models and Spectral Analysis for Sensory Time Series Clustering, by Jie Yin and Qiang Yang
- (R) Kernel-Density-Based Clustering of Time Series Subsequences Using a Continuous Random-Walk Noise Model, by Anne Denton
- (R) Efficient Query Filtering for Streaming Time Series, by Li Wei, Eamonn Keogh, Helga Van Herle, and Agenor Mafra-Neto
- (S) On the Stationarity of Multivariate Time Series for Correlation-Based Data Analysis, by Kiyoungh Yang and Cyrus Shahabi

- (S) Partial Elastic Matching of Time Series, by Longin Jan Latecki, Vasileios Megalooikonomou, Qiang Wang, Rolf Lakaemper, Chotirat Ann Ratanamahatana, and Eamonn Keogh
- (S) Mining Patterns That Respond to Actions, by Yuelong Jiang, Ke Wang, Alexander Tuzhilin, and Ada Wai-Chee Fu

Paper Session 2: Clustering Schemes I (Discovery Center A)

(3 regular papers-30 minutes each, and 6 short papers-10 minutes each)

Session Chair: Krishna Kummmamuru (kkummmamu@in.ibm.com)

- (R) Combining Multiple Clustering by Soft Correspondence, by Bo Long, Zhongfei (Mark) Zhang, and Philip S. Yu
- (R) Efficient Text Classification by Weighted Proximal SVM, by Dong Zhuang, Benyu Zhang, Qiang Yang, and Zheng Chen
- (R) A Framework of Labeling Unclustered Categorical Data into Clusters Based on the Important Attribute Values, by Hung-Leng Chen, Kun-Ta Chuang, and Ming-Syan Chen
- (S) Gradual Model Generator for Single-pass Clustering, by Ismo Kärkkäinen and Pasi Fränti
- (S) Bagging with Adaptive Costs, by Yi Zhang and Nick Street
- (S) On Feature Selection through Clustering, by Richard Butterworth, Gregory Piatetsky-Shapiro, and Dan Simovici
- (S) Adaptive Clustering: Obtaining Better Clusters Using Feedback and Past Experience, by Abraham Bagherjeiran, Christoph Eick, Chun-Sheng Chen, and Ricardo Vilalta
- (S) Mining Quantitative Frequent Itemsets Using Adaptive Density-based Subspace Clustering, by Takashi Washio, Yuki Mitsunaga, and Hiroshi Motoda
- (S) CLUMP: A Scalable and Robust Framework for Structure Discovery, by Kunal Punera and Joydeep Ghosh

Paper Session 3: Pattern Analysis on Text and Structured Data (Discovery Center B)

(3 regular papers-30 minutes each, and 6 short papers-10 minutes each)

Session Chair: Dino Pedreschi (pedre@di.unipi.it)

- (R) Mining Minimal Distinguishing Subsequence Patterns with Gap Constraints, by Xiaonan Ji, James Bailey, and Guozhu Dong
- (R) Neighborhood Formation and Anomaly Detection in Bipartite Graph, by Jimeng Sun, Huiming Qu, Deepayan Chakrabarti, and Christos Faloutsos
- (R) Shortest-path Kernels on Graphs, by Karsten Borgwardt and Hans-Peter Kriegel
- (S) An Optimal Linear Time Algorithm for Quasi-Monotonic Segmentation, by Daniel Lemire, Martin Brooks, and Yuhong Yan
- (S) Efficiently Mining Frequent Closed Partial Orders, by Jian Pei, Jian Liu, Haixun Wang, Ke Wang, Philip S. Yu, and Jianyong Wang
- (S) Mining Ontological Knowledge from Domain-Specific Text Documents, by Xing Jiang and Ah-Hwee Tan
- (S) Categorization and Keyword Identification of Unlabeled Documents, by Ning Kang, Carlotta Domeniconi, and Daniel Barbara

- (S) Fast Frequent String Mining Using Suffix Arrays, by Johannes Fischer, Volker Heun, and Stefan Kramer
- (S) Instability of Classifiers on Categorical Data, by Arno Siebes, Muhammad Subianto, and Ad Feelders

1:00 pm - 2:00 pm Lunch

2:00 pm -3:00 pm Keynote Speech (Legends I,II,III,IV)

Top 10 Data Mining Mistakes

Dr. John F. Elder IV (Elder Research, Inc., USA)

Data Mining is still as much it is an art as a science, and fancy new tools make it easy to do wrong things with one's data even faster. We'll examine the major "cracks in the crystal ball" through case studies, both simple and complex, of (often personal) errors t - drawn from real-world consulting engagements. Best Practices for Data Mining will be (accidentally) illuminated by their (rarely described) opposites. These common errors range from allowing anachronistic variables into the pool of candidate inputs, to subtly inflating results through early up-sampling. You'll hear cautionary tales of endangered projects and embarrassed teams - but also the keys to avoiding such a fate yourself.

3:00 pm - 3:30 pm Coffee Break

3:30 pm - 6:00 pm Paper Sessions (3 parallel Tracks)

Paper Session 4: Time Series II (Champions I,II)

(3 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Vasant Honavar (honavar@cs.iastate.edu)

- (R) An Algorithm for In-Core Frequent Itemset Mining on Streaming Data, by Ruoming Jin and Gagan Agrawal
- (R) Finding the Most Unusual Time Series Subsequence: Algorithms and Applications, by Eamonn Keogh, Jessica Lin, and Ada Fu
- (R) Finding Maximal Frequent Itemsets over Online Data Streams Adaptively, by Dae-su Lee and Wonsuk Lee
- (S) Feature Selection for Building Cost-Effective Data Stream Classifiers, by Like Gao and X. Sean Wang
- (S) Sequential Pattern Mining in Multiple Data Streams, by Gong Chen, Xindong Wu, and Xingquan Zhu
- (S) Mining Approximate Frequent Itemset from Noisy Data, by Jinze Liu, Paulsen Susan, Wei Wang, Andrew Nobel, and Jan Prins

Paper Session 5: Clustering Schemes II (Discovery Center A)

(4 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: George Kollios (gkollios@cs.bu.edu)

- (R) Effective and Efficient Distributed Model-based Clustering, by Hans-Peter Kriegel, Peer Kröger, Alexey Pryakhin, and Matthias Schubert

- (R) Making Subsequence Time Series Clustering Meaningful, by Jason Chen
- (R) A Generic Framework for Efficient Subspace Clustering of High-Dimensional Data, by Hans-Peter Kriegel, Peer Krö, Matthias Renz, and Sebastian Wurst
- (R) Online Hierarchical Clustering in a Data Warehouse Environment, by Elke Achtert, Christian Böans-Peter Kriegel, and Peer Krö
- (S) Hierarchical Density-Based Clustering of Uncertain Data, by Hans-Peter Kriegel and Martin Pfeifle
- (S) A Scalable Collaborative Filtering Framework based on Co-clustering, by Thomas George and Srujana Merugu
- (S) A Levelwise Search Algorithm for Interesting Subspace Clusters, by Haiyun Bian and Raj Bhatnagar

Paper Session 6: Quality Assessment (Discovery Center B)

(4 regular papers-30 minutes each)

Session Chair: Ryan Benton (rbenton@cacs.louisiana.edu)

- (R) Using Information-Theoretic Measures to Assess Association Rule Interestingness, by Julien Blanchard, Fabrice Guillet, Regis Gras, and Henri Briand
- (R) Ranking-Based Evaluation of Regression Models, by Saharon Rosset, Claudia Perlich, and Bianca Zadrozny
- (R) Discriminant Analysis: A Unified Approach, by Peng Zhang, Jing Peng, and Norbert Riedel
- (R) The Parameterized Complexity of Enumerating All, All Maximal, and All Closed Frequent Itemsets, by Matthew Hamilton, Rhonda Chaytor, and Todd Wareham

ICDM 2005 Tuesday, November 29, 2005

9:00 am - 10:00 am Keynote Speech (Legends I,II,III,IV)

Graphical Models for Structure Extraction and Information Integration

Dr. Sunita Sarawagi (IIT Bombay, India)

Recent advances in supervised learning over multiple inter-dependent variables have paved the way for accurate and automated methods for information extraction and integration. We present various graphical models for extraction, starting from traditional chain models for plain text, to segmentation models for exploiting matches with existing entities, and general graph models for extracting from visual 2D layouts as in web pages. Such models are trained either via conditional likelihood maximization or margin maximization leading to constrained convex optimization problems.

Inferencing often involves more than a simple message passing algorithm because of the presence of constraints that are not captured in the dependency graph. We present algorithms for such constrained inferencing and optimization tricks for reducing the computation of expensive features, like matches with large external dictionaries.

There is much scope for further research in handling diverse unstructured sources, continuous model refinement, efficient training and inferencing, and, probabilistic query answering in the presence of source uncertainties.

10:00 am - 10:15 am Coffee Break

10:15 am - 12:45 pm Paper Sessions (3 parallel Tracks)

Paper Session 7: Time Series III (Champions I,II)

(5 regular papers-30 minutes each)

Session Chair: Philip Chan (pkc@cs.fit.edu)

- (R) Mining Frequent Spatio-Temporal Sequential Patterns, by Huiping Cao
- (R) Extracting Frequent Subsequences from a Single Long Data Sequence: A Novel Anti-Monotonic Measure and a Simple On-Line Algorithm, by Koji Iwanuma, Ryuichi Ishihara, Yo Takano, and Hidetomo Nabeshima
- (R) Discriminatively Trained Markov Model for Sequence Classification, by Oksana Yakhnenko, Adrian Silvescu, and Vasant Honavar
- (R) WARP: Time Warping for Periodicity Detection, by Mohamed Elfeky, Walid Aref, and Ahmed Elmagarmid
- (R) Discovering Frequent Arrangements of Temporal Intervals, by Panagiotis Papapetrou, George Kollios, Stan Sclaroff, and Dimitrios Gunopulos

Paper Session 8: Spatial Data and Classification Schemes (Discovery Center A)

(4 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Frans Coenen (F.Coenen@csc.liv.ac.uk)

- (R) Compound Classification Models for Recommender Systems, by Lars Schmidt-Thieme
- (R) Multi-Stage Classification, by Ted Senator
- (R) Sharing Classifiers among Ensembles from Related Problem Domains, by Yi Zhang, Nick Street, and Samuel Burer
- (R) Parameter-Free Spatial Data Mining Using MDL, by Spiros Papadimitriou, Aristides Gionis, Panayiotis Tsaparas, Heikki Mannila, and Christos Faloutsos
- (S) Spatial Clustering Of Chimpanzee Locations For Neighborhood Identification, by Sandeep Mane, Carson Murray, Shashi Shekhar, Jaideep Srivastava, and Anne Pusey
- (S) A A Join-less Approach for Co-location Pattern Mining: A Summary of Results, by Jin Soung Yoo, Shashi Shekhar, and Mete Celik
- (S) A Graph-Based Ranking Algorithm for Geo-Referencing Documents, by Bruno Martins and Máo Silva

Paper Session 9:Preprocessing Techniques and Feature Selection (Discovery Center B)

(4 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Carlotta Domeniconi (carlotta@ise.gmu.edu)

- (R) Summarization - Compressing Data into an Informative Representation, by Varun Chandola and Vipin Kumar
- (R) A Bernoulli Relational Model for Nonlinear Embedding, by Gang Wang and Frederick Lochovsky
- (R) Stability of Feature Selection Algorithms, by Alexandros Kalousis, Julien Prados, and Melanie Hilario
- (R) Hierarchy-Regularized Latent Semantic Indexing, by Yi Huang, Kai Yu, Matthias Schubert, Shipeng Yu, and Hans-Peter Kriegel
- (S) Bit Reduction Support Vector Machine, by Lawrence Hall, Tong Luo, Dmitry Goldof, and Andrew Remsen
- (S) Speculative Markov Blanket Discovery for Optimal Feature Selection, by Sandeep Yaramakala and Dimitris Margaritis
- (S) Bias Analysis in Text Classification for Highly Skewed Data, by Lei Tang and Huan Liu

12:45 pm - 2:15 pm Lunch

2:15 pm - 3:15 pm Panel Session (Legends I,II,III,IV)

Data mining, where to go?

Organizer: Wen-Ran Zhang (Georgia Southern University, USA)

Panelists:

Jaiwei Han, University of Illinois at Urbana-Champaign, USA. Topic: “Exploring New Applications.”

Vijay Raghavan, University of Louisiana, Lafayette, USA. Topic: “Web Content Mining.”

Bamshad Mobasher, DePaul University, Chicago, USA. Topic: “Personalization and User Modeling.”

Ramamohanarao Kotagiri, University of Melbourne, Australia. Topic: “Data Mining & Machine Learning.”

Wen-Ran Zhang, Georgia Southern University, USA. Topic: “Multiagent Data Warehousing (MADWH) and Multiagent Data Mining (MADM).”

3:15 pm - 3:30 pm Coffee Break

3:30 pm - 6:00 pm Paper Sessions (3 parallel Tracks)

Paper Session 10: Learning Techniques I (Champions I,II)

(4 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Wei Fan (weifan@us.ibm.com)

- (R) Supervised Tensor Learning, by Dacheng Tao, Xuelong Li, Weiming Hu, Stephen Maybank, and Xindong Wu
- (R) Learning Instance Greedily Cloning Naive Bayes for Ranking, by Liangxiao Jiang and Harry Zhang
- (R) Improving Automatic Query Classification via Semi-supervised Learning, by Steven Beitzel, Eric Jensen, David Lewis, Abdur Chowdhury, Aleksander Kolcz, and Ophir Frieder

- (R) X-mHMM: An Efficient Algorithm for Training Mixtures of HMMs when the Number of Mixtures is Unknown, by ZoltáSzamonek and Csaba Szepesvá
- (S) Supervised Ordering --- An Empirical Survey, by Toshihiro Kamishima, Hideto Kazawa, and Shotaro Akaho
- (S) A Framework for Semi-Supervised Learning based on Subjective and Objective Clustering Criteria, by Maria Halkidi, Dimitrios Gunopulos, Nitin Kumar, Michalis Vazirgiannis, and Carlotta Domeniconi
- (S) A Preference Model for Structured Supervised Learning Tasks, by Fabio Aiolli

Paper Session 11: Data Representation (Discovery Center A)

(4 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Haesun Park (hpark@cc.gatech.edu)

- (R) Finding Representative Set from Massive Data, by Feng Pan, Wei Wang, Anthony K. H. Tung, and Jiong Yang
- (R) eMailSift: Email Classification Based on Structure and Content, by Manu Aery and Sharma Chakravarthy
- (R) Orthogonal Neighborhood Preserving Projections, by Effrosyni Kokiopoulou and Yousef Saad
- (R) A Heterogeneous Field Matching Method for Record Linkage, by Steven Minton, Claude Nanjo, Craig Knoblock, Martin Michalowski, and Matthew Michelson
- (S) Text Classification with Evolving Label-sets, by Shantanu Godbole, Ganesh Ramakrishnan, and Sunita Sarawagi
- (S) Efficient Mining of High Branching Factor Attribute Trees, by Alexandre Termier, Marie-Christine Rousset, Michele Sebag, Kouzou Ohara, Takashi Washio, and Hiroshi Motoda
- (S) Text Representation and Dimension Reduction: from Vector to Tensor, by Ning Liu, Jun Yan, Benyu Zhang, Zheng Chen, Fengshan Bai, and Qiansheng Cheng

Paper Session 12: Security and Privacy (Discovery Center B)

(3 regular papers-30 minutes each, and 5 short papers-10 minutes each)

Session Chair: Christopher W. Clifton (clifton@cs.purdue.edu)

- (R) Approximate Inverse Frequent Itemset Mining: Privacy, Complexity, and Approximation, by Yongge Wang and Xintao Wu
- (R) A Border-Based Approach for Hiding Sensitive Frequent Itemsets, by Xingzhi Sun and Philip S. Yu
- (R) Template-Based Privacy Preservation in Classification Problems, by Ke Wang, Benjamin C. M. Fung, and Philip S. Yu
- (S) Privacy-Preserving Frequent Pattern Mining across Private Databases, by Ada Wai-Chee Fu, Raymond Chi-Wing Wong, and Ke Wang
- (S) Suppressing Data Sets to Prevent Discovery of Association Rules, by Ayca Azgin Hintoglu, Ali Inan, Yucel Saygin, and Mehmet Keskinöz
- (S) Blocking Anonymity Threats Raised by Frequent Itemset Mining, by Maurizio Atzori, Francesco Bonchi, Fosca Giannotti, and Dino Pedreschi
- (S) A Random Rotation Perturbation Approach to Privacy Preserving Data Classification, by Keke Chen and Ling Liu

- (S) Segment-Based Injection Attacks against Collaborative Recommender Systems, by Robin Burke, Bamshad Mobasher, Runa Bhaumik, and Chad Williams

7:15 pm Banquet at NASA Houston Space Center

ICDM 2005 Wednesday, November 30, 2005

9:00 am - 10:00 am Keynote Speech (Legends I,II,III,IV)

Efficient Indexing Technology for Data Mining of Scientific Data

Arie Shoshani, Lawrence Berkeley National Laboratory, USA

Data mining in scientific applications usually involves searches over a large number of objects in the multidimensional space of their properties, or searches for known patterns. This is in contrast to mining for associations between objects, or discovering new patterns. Examples are searching over billions of objects to find rare objects by expressing numerical range conditions on their properties, or finding flame fronts in large volume, spatio-temporal combustion simulation data by expressing multiple conditions over the data values associated with the cells in the 3D space. A critical issue in supporting such directed searches over large data volumes is the efficiency of the indexing method. This is required in order to facilitate real time exploration of the data. In this talk, we will describe a specialized bitmap indexing method, called FastBit, which has proved especially appropriate for numeric multidimensional data common in scientific applications. We will illustrate the use of this technology with several examples.

10:00 am - 10:30 am Coffee Break

10:30 am - 1:00 pm Paper Sessions (3 parallel Tracks)

Paper Session 13: Learning Techniques II (Champions I,II)

(4 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Xintao Wu (xwu@uncc.edu)

- (R) Training Support Vector Machines using Gilbert's Algorithm, by Shawn Martin
- (R) CanTree: A Tree Structure for Efficient Incremental Mining of Frequent Patterns, by Carson Kai-Sang Leung, Quamrul I. Khan, and Tariqul Hoque
- (R) Balancing Exploration and Exploitation: A New Algorithm for Active Machine Learning, by Thomas Osugi, Deng Kun, and Stephen Scott
- (R) Classifier Fusion Using Shared Sampling Distribution for Boosting, by Costin Barbu, Raja Iqbal, and Jing Peng
- (S) Obtaining Best Parameter Value for Accurate Classification, by Frans Coenen and Paul Leng
- (S) CloseMiner: Discovering Frequent Closed Itemsets using Frequent Closed Tidsets, by Gourakishwar Ningthoujam, Ranbir Sanasam, and Anjana Kakoti

- (S) Semi-Supervised Clustering with Metric Learning using Relative Comparisons, by Nimit Kumar, Krishna Kummamuru, and Deepa Paranjpe

Paper Session 14: Data Mining Applications: Bio-Medical and Social (Discovery Center A)

(3 regular papers-30 minutes each, and 6 short papers-10 minutes each) Session Chair: Sunita Sarawagi (sunita@it.iitb.ac.in)

- (R) Alternate Representation of Distance Matrices for Characterization of Protein Structure, by Keith Marsolo and Srinivasan Parthasarathy
- (R) SVM Feature Selection for Classification of SPECT Images of Alzheimer's Disease using Spatial Information, by Glenn Fung and Jonathan Stoeckel
- (R) ViVo: Visual Vocabulary Construction for Mining Biomedical Images, by Arnab Bhattacharya, Vebjorn Ljosa, Jia-Yu Pan, Mark Verardo, Hyunjeong Yang, Christos Faloutsos, and Ambuj K. Singh
- (S) CLUGO: A Clustering Algorithm for Automated Functional Annotations Based on Gene Ontology, by In-Yee Lee, Jan-Ming Ho, and Ming-Syan Chen
- (S) A Cooperative Data Mining Approach and Its Application to Early Diabetes Detection, by Jie Gao, Joerg Denzinger, and Robert C. James
- (S) Face Recognition Using Landmark-based Bidimensional Regression, by Jiazheng Shi, Ashok Samal, and David Marx
- (S) A Computational Framework for Taxonomic Research: Diagnosing Body Shape within Fish Species Complexes, by Yixin Chen, Henry Bart, Shuqing Huang, and Huimin Chen
- (S) Focused Community Discovery, by Kirsten Hildrum and Philip Yu
- (S) Pruning Social Networks Using Structural Properties and Descriptive Attributes, by Lisa Singh, Lise Getoor, and Louis Licamele

Paper Session 15: Statistical Methods I (Discovery Center B)

(4 regular papers-30 minutes each, and 2 short papers-10 minutes each)
Session Chair: Ramamohanarao Kotagiri (rao@csse.unimelb.edu.au)

- (R) Generalizing the Notion of Confidence, by Michael Steinbach and Vipin Kumar
- (R) A new algorithm for finding Minimal Sample Uniques for use in Statistical Disclosure Assessment, by Anna Manning and David Haglin
- (R) Effective Estimation of Posterior Probabilities: Explaining the Accuracy of Randomized Decision Tree Approaches, by Wei Fan, Ed Greengrass, Joe McClosky, Philp Yu, and Kevin Drummey
- (R) An Empirical Bayes Approach to Detect Anomalies in Dynamic Multidimensional Arrays, by Deepak Agarwal
- (S) Dynamic Ensemble Re-Construction for Better Ranking, by Jin Huang and Charles X. Ling
- (S) An Improved Categorization of Classifier's Sensitivity on Sample Selection Bias, by Wei Fan, Ian Davidson, Bianca Zadrozny, and Philip S. Yu

1.15 pm -2.00 pm ICDM Business Meeting (bring your conference lunch box with you!)

- 10 Challenging Problems in Data Mining, by Qiang Yang
- Data Mining on ICDM '05 Paper Submissions, by Shusaku Tsumoto
- ICDM '06 in Hong Kong, by Chris Clifton and Ning Zhong

2:00 pm - 3:30 pm Paper Sessions (3 parallel Tracks)

Paper Session 16: Learning Techniques II (Champions I,II)

(1 regular papers-30 minutes each, and 6 short papers-10 minutes each)

Session Chair: Haixun Wang (haixun@us.ibm.com)

- (R) Adaptive Product Normalization: Using Online Learning for Record Linkage in Comparison Shopping, by Mikhail Bilenko, Sugato Basu, and Mehran Sahami
- (S) A Rule Evaluation Support Method with Learning Models Based on, by Hidenao Abe, Shusaku Tsumoto, Miho Ohsaki, and Takahira Yamaguchi
- (S) Learning through Changes: An Empirical Study of Dynamic Behaviors of Probability Estimation Trees, by Kun Zhang, Zujia Xu, Jing Peng, and Bill Buckles
- (S) On Learning Asymmetric Dissimilarity Measures, by Krishna Kummamuru, Raghu Krishnapuram, and Rakesh Agrawal
- (S) Semi-Supervised Mixture of Kernels via LPBoost Methods, by Jinbo Bi, Glenn Fung, Murat Dundar, and Bharat Rao
- (S) Mining Chains of Relations, by Aristides Gionis, Foto Afrati, Gautam Das, Heikki Mannila, Taneli Mielikainen, and Panayiotis Tsaparas
- (S) Anomaly Intrusion Detection using Multi-Objective Genetic Fuzzy System and Agent-based Evolutionary Computation Framework, by Chi-Ho Tsang, Sam Kwong, and Hanli Wang

Paper Session 17: Data Mining Applications: Web (Discovery Center A)

(2 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Christoph Eick (ceick@uh.edu)

- (R) Usage-based PageRank for Web Personalization, by Magdalini Eirinaki and Michalis Vazirgiannis
- (R) Higher-Order Web Link Analysis Using Multilinear Algebra, by Tamara Kolda, Brett Bader, and Joseph Kenny
- (S) Merging Interface Schemas on the Deep Web via Clustering Aggregation, by Wensheng Wu, AnHai Doan, and Clement Yu
- (S) Automatically Mining Result Records from Search Engine Response Pages, by Dheerendranath Mundluru
- (S) Hot Item Mining and Summarization from Multiple Auction Web Sites, by Tak-Lam Wong and Wai Lam

Paper Session 18: Statistical Methods II (Discovery Center B)

(2 regular papers-30 minutes each, and 5 short papers-10 minutes each)

Session Chair: Martin Scholz (scholz@kimo.cs.uni-dortmund.de)

- (R) Leveraging Relational Autocorrelation with Latent Group Models, by Jennifer Neville and David Jensen
- (R) A Random Walk through Human Associations, by Raz Tamir
- (S) Triple Jump Acceleration for the EM Algorithm, by Han-Shen Huang, Chun-Nan Hsu, and Bou-Ho Yang
- (S) Economical Active Feature-value Acquisition through Expected Utility Estimation, by Prem Melville, Maytal Saar-Tsechansky, Foster Provost, and Raymond Mooney
- (S) Example-Based Robust Outlier Detection in High Dimensional Datasets, by Cui Zhu, Hiroyuki Kitagawa, and Christos Faloutsos
- (S) Pairwise Symmetry Decomposition Method for Generalized Covariance Analysis, by Tsuyoshi Ide
- (S) FS3: A Random Walk based Free-Form Spatial Scan Statistic for Anomalous Window Detection, by Vandana Janeja and Vijayalakshmi Atluri

3:00 pm - 4:00 pm Coffee Break

4:00 pm - 5:30 pm Paper Sessions (3 parallel Tracks)

Paper Session 19: Tools and Algorithms (Champions I, II)

(1 regular papers-30 minutes each, and 6 short papers-10 minutes each)

Session Chair: Gautam Das (gdas@cse.uta.edu)

- (R) A Visual Data Mining Framework for Convenient Identification of Useful Knowledge, by Kaidi Zhao, Bing Liu, Thomas Tirpak, and Weimin Xiao
- (S) Visualizing Global Manifold Based on Distributed Local Data Abstraction, by Xiaofeng Zhang and William K. Cheung
- (S) Making Logistic Regression A Core Data Mining Tool, by Paul Komarek and Andrew Moore
- (S) Parallel Algorithms for distance-based and density-based outliers, by Elio Lozano and Edgar Acuna
- (S) Optimizing Constraint-Based Mining by Automatically Relaxing Constraints, by Arnaud Soulet and Bruno Cr elleux
- (S) CTC - Correlating Tree Patterns for Classification, by Albrecht Zimmermann and Bjoern Bringmann
- (S) On the Complexity of Rule Discovery from Distributed Data, by Martin Scholz

Paper Session 20: Data Mining Applications (Discovery Center A)

(2 regular papers-30 minutes each, and 3 short papers-10 minutes each)

Session Chair: Hiroyuki Kawano (kawano@it.nanzan-u.ac.jp)

- (R) AMIOT: Induced Ordered Tree Mining in Tree-structured Databases, by Shohei Hido and Hiroyuki Kawano
- (R) Mining Patterns of Change in Remote Sensing Image Databases, by Marcelino Pereira S. Silva, Gilberto C ara, Ricardo Cartaxo M. Souza, Dalton M. Valeriano, and Maria Isabel S. Escada

- (S) Process Diagnosis via Electrical-Wafer-Sorting Maps Classification, by Federico Di Palma, Giuseppe De Nicolao, and Guido Miraglia
- (S) Average Number of Frequent (Closed) Patterns in Bernoulli and Markovian Databases, by Loick Lhote, Francois Rioult, and Arnaud Soulet
- (S) Predicting Software Escalations with Maximum ROI, by Charles X. Ling, Shengli Sheng, Tilmann Bruckhaus, and Nazim H. Madhavji

Paper Session 21: Optimization Techniques (Discovery Center B)

(3 regular papers-30 minutes each)

Session Chair: Mohammed El-Hajj (mohammad@cs.ualberta.ca)

- (R) Handling Generalized Cost Functions in the Partitioning Optimization Problem Through Sequential Binary Programming, by Alan Abrahams, Adrian Becker, Daniel Fleder, and Ian MacMillan
- (R) Bifold Constraint-Based Mining by Simultaneous Monotone and Anti-Monotone Checking, by Mohammad El-Hajj, Osmar Zaiane, and Paul Nalos
- (R) A Thorough Experimental Study of Datasets for Frequent Itemsets, by Frédéric Flouvat, Fabien De Marchi, and Jean-Marc Petit