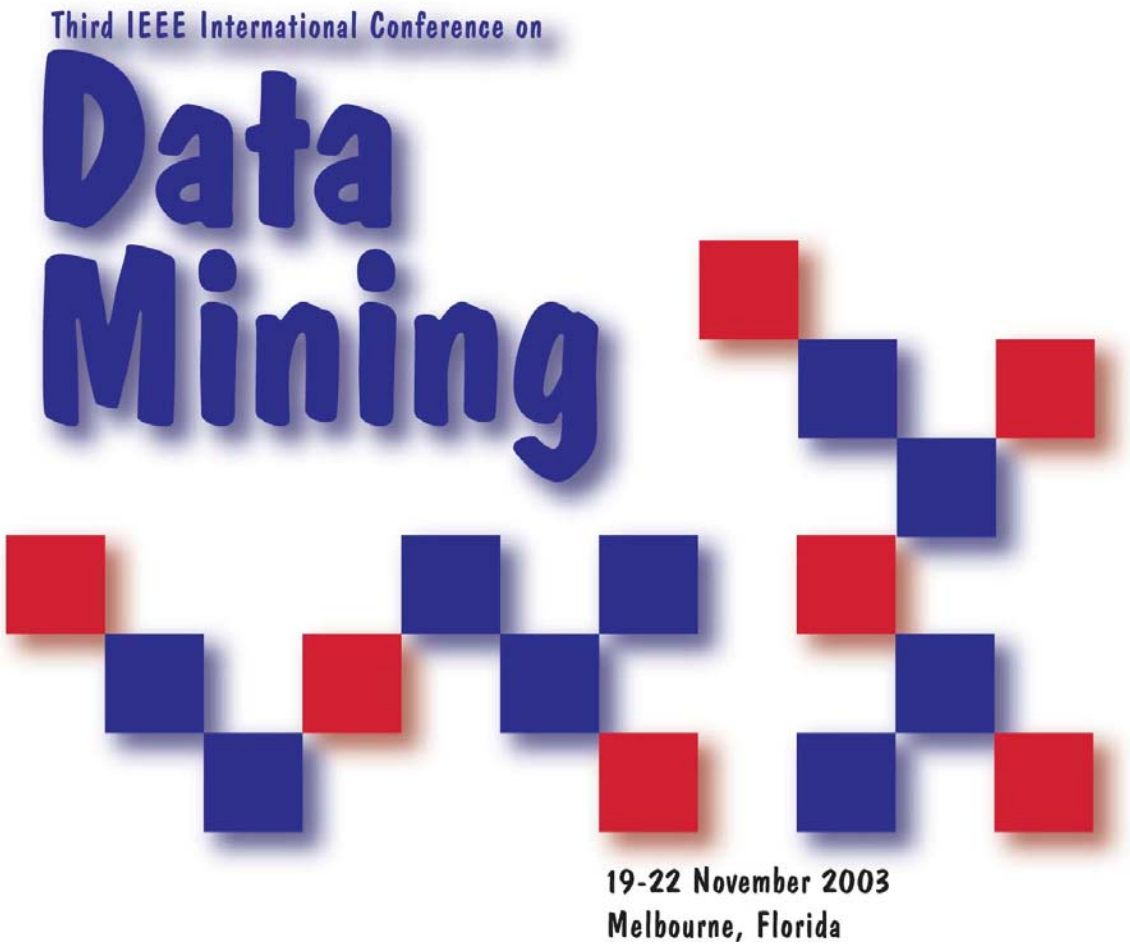

Conference Schedule

IEEE International Conference on Data Mining (ICDM 2003)

Melbourne, Florida USA

November 19-22, 2003

The Conference at a Glance

Each morning (8-9am Wednesday; 7:30-8:30am, Thursday-Saturday) a continental breakfast will be available on the Terrace and in the back of Barbados. It is included in the registration fee.

The registration desk will be open during at least the following hours:

Wednesday:	8-9am, 10:30am-2pm, and 6-9pm	Friday:	7:30-8:30am and 9:40-11am
Thursday:	7:30-8:30am and 10am-1:30pm	Saturday:	7:30-8:30am

The last page of this schedule contains a figure showing the hotel's floor plan. The numbers in parentheses – e.g., **(1a)** – are the identifiers for the technical sessions.

The schedule for the workshops and tutorials on Wednesday November 19 appears on page 4.

ICDM 2003 – Thursday (November 20)

8:30-8:50	**Grand Caribbean** Welcome and Introductory Remarks			
8:50-10:00	**Grand Caribbean** Invited Talk	***Pattern Discovery for Genomics*** Gene Myers		
10:00-10:30	**Break**	**Break**	**Break**	**Break**
10:30-12:00	**Barbados (1a)** Clustering 1	**Aruba (1b)** Visualization 1	**Trinidad (1c)** Mining User Behavior 1	**Penthouse (1d)** Issues in Supervised Learning
12:00-1:30	**Lunch on Oceanfront Deck** (included in registration)			
1:30-2:40	**Grand Caribbean** Invited Talk	***Real-Time Monitoring and Surveillance using Data Stream Mining*** Philip Yu		
2:45-3:40	**Barbados (2a)** Clustering 2	**Aruba (2b)** Applications Track 1	**Trinidad (2c)** Rule-Based Methods	**Penthouse (2d)** Tree-Based Methods
3:40-4:10	**Break**	**Break**	**Break**	**Break**
4:10-5:40	**Barbados (3a)** Text 1	**Aruba (3b)** Spatial and Temporal Tasks	**Trinidad (3c)** Part A: Mining User Behavior 2 Part B: Web Data Mining	**Penthouse (3d)** Part A: Bayesian Networks Part B: Ensembles
7:00-10:30	**Poster Session and Reception in Grand Caribbean Ballroom**			

ICDM 2003 – Friday (November 21)

	Grand Caribbean	***Grand Challenges on the Road to Practical Data Mining Systems***		
8:30-9:40	Invited Talk	Usama Fayyad		
9:40-10:10	**Break**	**Break**	**Break**	**Break**
	Barbados (4a)	**Aruba (4b)**	**Trinidad (4c)**	**Penthouse (4d)**
10:10-11:40	Association Rules 1	Text 2	Data Cleaning and Methodological Issues	Feature Selection
11:40-12:00	**Box Lunch on Oceanfront Deck** (included in registration)			
12:00-6:30	**Guided Tour of NASA Kennedy Space Center** (included in registration)			
6:30-10:00	**Conference Banquet in Grand Caribbean Ballroom**			

ICDM 2003 – Saturday (November 22)

	Grand Caribbean	***Sequential Supervised Learning: Methods for Sequence Labeling and Segmentation***		
8:30-9:40	Invited Talk	Thomas Dietterich		
9:40-10:10	**Break**	**Break**	**Break**	**Break**
	Barbados (5a)	**Aruba (5b)**	**Trinidad (5c)**	**Penthouse (5d)**
10:10-12:00	Part A: Privacy-Preserving Data Mining Part B: Databases and Data Mining	Mining Sequential and Hierarchical Data	Mining Frequent Items	Clustering 3
12:00-1:30	**Lunch** (included in registration) **on Oceanfront Deck**			
12:45-1:30	**TCCI Business Meeting in Grand Caribbean Ballroom**			
	Grand Caribbean	***Global Structure from Sequences***		
1:30-2:40	Invited Talk	Heikki Mannila		
	Barbados		**Trinidad (6)**	
2:45-3:45	Panel on **Security and Data Mining: Funding Priorities and Opportunities** (Michael Pazzani, Chair)		Support Vector Machines and Nearest-Neighbor Methods	
3:45-4:15	**Break**	**Break**	**Break**	**Break**
	Barbados (7a)	**Aruba (7b)**	**Trinidad (7c)**	**Penthouse (7d)**
4:15-5:45	Linkage-Based Methods	Part A: Applications Track 2 Part B: Bioinformatics	Association Rules 2	Visualization 2 and Image Processing

ICDM 2003 Wednesday (November 19, 2003)

There will a 30min break in the morning break at 10:30am and in the afternoon at 3:30pm. Lunch will be 12:30-2pm on the Oceanfront Deck and is included for workshop and tutorial attendees. Also included in the registration fee is the continental breakfast from 8-9am; it will be available on the Terrace and in the back of Barbados.

Tutorials

Morning (9am – 12:30am)

Trinidad/Antigua	*Bioinformatics and Machine Learning Methods* given by Chris Ding
Aruba	*Information Extraction: Theory and Practice* given by Ronen Feldman

Afternoon (2pm – 5:30pm)

Trinidad/Antigua	*Advances in Clustering and Applications* given by Alexander Hinneburg and Daniel Keim
Aruba	*Data Mining for Security Applications* given by Aleksandar Lazarevic, Jaideep Srivastava, and Vipin Kumar

Workshops (9am – 5:30pm)

St. Anne	*Clustering Large Data Sets* organized by Daniel Boley, Inderjit Dhillon, Joydeep Ghosh, and Jacob Kogan
Martinique	*Data Mining for Computer Security (DMSEC '03)* organized by Philip Chan, Vipin Kumar, Wenke Lee, and Srinivasan Parthasarathy
Barbados	*Foundations and New Directions in Data Mining* organized by T. Y. Lin (co-chair), S. Ohsuga (co-chair), Tony Hu, and C. J. Liau
Penthouse	*Frequent Itemset Mining Implementations (FIMI '03)* organized by Bart Goethals and Mohammed J. Zaki
St. Thomas	*Privacy Preserving Data Mining (PPDM)* organized by Wenliang (Kevin) Du (chair) and Chris Clifton (co-chair)
St. Croix	*VDM@ICDM2003: The 3rd International Workshop on Visual Data Mining* organized by Simeon Simoff, Monique Noirhomme-Fraiture, and Michael Bvhlen

ICDM 2003 Thursday (November 20, 2003)

8:30-8:50am Welcome & Introductory Remarks (Grand Caribbean)

8:50-10:00am Keynote Address (Grand Caribbean; intro'ed by Jude Shavlik)

Pattern Discovery for Genomics

Gene Myers, University of California, Berkeley, USA

We now have the sequence of seven animal genomes with the prospect of having one to two hundred within the next few years. The next great challenge in genomics is to decipher what is coded within these sequences. This talk will survey the state of the art in finding genes and regulatory signals with an emphasis on the first attempts at using cross-species comparisons to do so.

10:00-10:30am Coffee Break

10:30-12:00pm Technical Sessions 1

Regular papers are allocated 20 minutes for presentation and 2 minutes for questions. Short papers are allocated 10 minutes for presentation and 1 minute for questions.

Barbados Room – Clustering 1 (Session 1a Chair: David Page)

10:30am (22 min)	*Localized prediction of continuous target variables using hierarchical clustering,* by Aleksandar Lazarevic, Ramdev Kanapady, Chandrika Kamath, Vipin Kumar, and Kumar Tamma
10:52am (11min)	*Class decomposition via clustering: A new framework for low-variance classifiers*, by Ricardo Vilalta, Murali-Krishna Achari, and Christoph Eick
11:03am (22 min)	*Model-based clustering with soft balancing,* by Shi Zhong and Joydeep Ghosh
11:25am (11 min)	*Fast PNN-based clustering using k-nearest neighbor graph,* by Pasi Frdnti, Olli Virmajoki, and Ville Hautamdki
11:36am (22 min)	*OP-Cluster: Clustering by tendency in high dimensional space*, by Jinze Liu and Wei Wang

Aruba Room – Visualization 1 (Session 1b Chair: Einoshin Suzuki)

10:30am (22 min)	*Visualization of rule's similarity using multidimensional scaling,* by Shusaku Tsumoto and S. Hirano
10:52am (11min)	*Icon-based visualization of large high-dimensional datasets,* by Ping Chen, Chenyi Hu, Wei Ding, Heloise Lynn, and Yves Simon
11:03am (22 min)	*Interactive visualization and navigation in large data collections using the hyperbolic space*, by Jörg Walter, Jörg Ontrup, Daniel Wessling, and Helge Ritter
11:25am (11 min)	*A user-driven and quality-oriented visualization for mining association rules*, by Julien Blanchard, Fabrice Guillet, and Henri Briand
11:36am (11 min)	*Validating and refining clusters via visual rendering*, by Keke Chen and Ling Liu
11:47am (11 min)	*Towards simple, easy-to-understand, yet accurate classifiers*, by Doina Caragea, Dianne Cook, and Vasant Honavar

Trinidad Room – Mining User Behavior 1 (Session 1c Chair: Osmar Zaiane)

10:30am (22 min) *Semantic log analysis based on a user's query behavior model*, by Noriaki Kawamae, T. Mukaigaito, and M. Hanaki

10:52am (22 min) *MPIS: Maximal-profit item selection with cross-selling considerations,* by Raymond Chi-Wing Wong, Ada Wai-Chee Fu, and Ke Wang

11:14am (22 min) *Probabilistic user behavior models,* by Eren Manavoglu, Dmitry Pavlov, and C. Lee Giles

11:36am (22 min) *Mining plans for customer-class transformation,* by Qiang Yang and Hong Cheng

Penthouse – Issues in Supervised Learning (Session 1d Chair: Zoran Obradovic)

10:30am (22 min) *Exploiting unlabeled data for improving accuracy of predictive data mining,* by Kang Peng, Slobodan Vucetic, Bo Han, Hongbo Xie, and Zoran Obradovic

10:52am (11min) *Using discriminant analysis for multi-class classification,* by Tao Li, S. Zhu, and M. Ogihara

11:03am (22 min) *Cost-sensitive learning by cost-proportionate example weighting,* by Bianca Zadrozny, John Langford, and Naoki Abe

11:25am (22 min) *A new optimization criterion for generalized discriminant analysis on undersampled problems,* by Jieping Ye, Ravi Janardan, Cheong Hee Park, and Haesun Park

12:00-1:30pm Lunch (on Oceanfront Deck; included in registration)

1:30-2:40pm Invited Talk (Grand Caribbean; introduced by Philip Chan)

Real-Time Monitoring and Surveillance using Data Stream Mining

Philip Yu, IBM T.J. Watson Research Center, USA

With the advance of data gathering and communication technologies, it becomes increasingly possible to support real-time monitoring of large amount of information from diverse information sources. Examples include trade surveillance for security fraud and money laundering, network monitoring for intrusion detection, bio-surveillance for terrorist attacks, and various sensor network based monitoring applications. Data is viewed as a continuous stream in these kinds of applications. Problems such as data mining which have been widely studied for traditional data sets cannot be easily applied to the data stream domain. This is because the large volume of data arriving in a stream renders most algorithms too inefficient as most mining algorithms require multiple scans of data which is unrealistic for stream data. More importantly, the characteristics of the data stream can change over time and the evolving pattern needs to be captured. In this talk, I'll provide an overview, discuss the issues and focus on how to mine evolving data streams.

2:45-3:40pm Technical Sessions 2

Barbados Room – Clustering 2 (Session 2a Chair: Wei Fan)

2:45pm (11 min)	*Facilitating fuzzy association rules mining by using multi-objective genetic algorithms for automated clustering,* by Mehmet Kaya and Reda Alhajj
2:56pm (11 min)	*Information theoretic clustering of sparse co-occurrence data,* by Inderjit Dhillon and Yuqiang Guan
3:07pm (11 min)	*Frequent-pattern based iterative projected clustering,* by Man Lung Yiu and Nikos Mamoulis
3:18pm (11 min)	*Improving home automation by discovering regularly occurring device usage patterns,* by Edward Heierman and Diane Cook
3:29pm (11 min)	*Clustering item data sets with association-taxonomy similarity,* by Ching-Huang Yun, Kun-Ta Chuang, and Ming-Syan Chen

Aruba Room – Applications Track 1 (Session 2b Chair: Tony Hu)

2:45pm (11 min)	*Inference of protein-protein interactions by unlikely profile pair,* by Byung-Hoon Park, George Ostrouchov, Gong-Xin Yu, Al Geist, Andrey Gorin, and Nagiza Samatova
2:56pm (11 min)	*Applying noise handling techniques to genomic data: A case study,* by Choh Man Teng
3:07pm (11 min)	*Predicting distribution of a new forest disease using one-class SVMs,* by Qinghua Guo, Maggi Kelly, and Catherine Graham
3:18pm (11 min)	*Understanding Helicoverpa armigera pest population dynamics related to chickpea crop using neural networks,* by Rajat Gupta, B. Narayana, Krishna Polepalli, G. Ranga Rao, C. Gowda, Y. Reddy, and G. Rama Murthy

Trinidad Room – Rule-Based Methods (Session 2c Chair: Taghi Khoshgoftaar)

2:45pm (22 min)	*Direct interesting rule generation,* by Jiuyong Li and Yanchun Zhang
3:07pm (11 min)	*Learning rules for anomaly detection of hostile network traffic,* by Matthew Mahoney and Philip Chan
3:18pm (11 min)	*Pattern discovery based on rule induction and taxonomy generation,* by Shusaku Tsumoto and S. Hirano
3:29pm (11 min)	*Bootstrapping rule induction,* by Lemuel Waitman, Douglas Fisher, and Paul King

Penthouse – Tree-Based Methods (Session 2d Chair: Charles Ling)

2:45pm (11 min)	*Indexing and mining free trees,* by Yun Chi, Yirong Yang, and Richard Muntz
2:56pm (11 min)	*T-trees, vertical partitioning and distributed association rule mining,* by Frans Coenen, Paul Leng, and Shakil Ahmed
3:07pm (11 min)	*Tree-structured partitioning based on splitting histograms of distances,* by Longin Jan Latecki, Rajagopal Venugopal, Marc Sobel, and Steve Horvath
3:18pm (11 min)	*Postprocessing decision trees to extract actionable knowledge,* by Qiang Yang, Jie Yin, Charles Ling, and Tielin Chen

3:40-4:10pm Coffee Break

4:10-5:40pm Technical Sessions 3

Barbados Room – Text 1 (Session 3a Chair: Philip Yu)

4:10pm (22 min)	*Statistical relational learning for document mining,* by Alexandrin Popescul, Lyle Ungar, Steve Lawrence, and David Pennock
4:32pm (11 min)	*Mining relevant text from unlabelled documents,* by Daniel Barbara, Carlotta Domeniconi, and Ning Kang
4:43pm (11 min)	*Ontologies improve text document clustering,* by Andreas Hotho, Steffen Staab, and Gerd Stumme
4:54pm (22 min)	*Building text classifiers using positive and unlabeled data,* by Bing Liu, Y. Dai, Xiaoli Li, Wee Sun Lee, and Philip Yu
5:16pm (11 min)	*Mining the web to discover the meanings of an ambiguous word*, by Raz Tamir and Reinhard Rapp
5:27pm (11 min)	*Semantic role parsing: Adding semantic structure to unstructured text,* by Sameer Pradhan, Kadri Hacioglu, Wayne Ward, James Martin, and Dan Jurafsky

Aruba Room – Spatial and Temporal Tasks (Session 3b Chair: Martin Ester)

4:10pm (22 min)	*Complex spatial relationships,* by Robert Munro, Sanjay Chawla, and Pei Sun
4:32pm (11 min)	*PixelMaps: A new visual data mining approach for analyzing large spatial data sets,* by Daniel Keim, Christian Panse, M. Sips, and Stephen North
4:43pm (11 min)	*Efficient subsequence matching in time series databases under time and amplitude transformations,* by Tassos Argyros and Charis Ermopoulos
4:54pm (22 min)	*Efficient multidimensional quantitative hypotheses generation,* by Amihood Amir, Reuven Kashi, and Nathan Netanyahu
5:16pm (22 min)	*Clustering of time series subsequences is meaningless: Implications for previous and future research,* by Eamonn Keogh, Jessica Lin, and Wagner Truppel

Trinidad Room – Part A: Mining User Behavior 2, Part B: Web Data Mining
(Session 3c Chair: Shusaku Tsumoto)

4:10pm (22 min)	*Segmenting customer transactions using a pattern-based clustering approach,* by Yinghui Yang and Balaji Padmanabhan
4:32pm (11 min)	*Combining the web content and usage mining to understand the visitor behavior in a web site,* by Juan Velasquez, Hiroshi Yasuda, and Terumasa Aoki
4:43pm (11 min)	*The hybrid Poisson aspect model for personalized shopping recommendation,* by Chun-Nan Hsu, Hao-Hsiang Chung, and Han-Shen Huang
4:54pm (22 min)	*Integrating customer value considerations into predictive modeling,* by Saharon Rosset and Einat Neumann
5:16pm (22 min)	*On precision and recall of multi-attribute data extraction from semistructured sources,* by Guizhen Yang, Saikat Mukherjee, and I. V. Ramakrishnan

Penthouse – Part A: Bayesian Networks, Part B: Ensembles
(Session 3d Chair: Choh Man Teng)

4:10pm (11 min)	*Simple estimators for relational Bayesian classifiers,* by Jennifer Neville, David Jensen, and Brian Gallagher
4:21pm (11 min)	*Structure search and stability enhancement of Bayesian networks,* by Hanchuan Peng and Chris Ding
4:32pm (22 min)	*Dynamic weighted majority: A new ensemble method for tracking concept drift,* by Jeremy Kolter and Marcus Maloof
4:54pm (11 min)	*Ensembles of cascading trees,* by Jinyan Li and Huiqing Liu
5:05pm (11 min)	*Comparing pure parallel ensemble creation techniques against bagging,* by Lawrence Hall, Kevin Bowyer, R. Banfield, D. Bhadoria, W. Kegelmeyer, and S. Eschrich

7:00-10:30pm Reception and Poster Session (Grand Caribbean)

First authors whose last names are in the range *A-L* are expected to staff their posters from 7-9pm. First authors in the range *M-Z* are expected to be at their posters from 8:30-10:30pm.

ICDM 2003 Friday (November 21, 2003)

8:30-9:40am Invited Talk (Grand Caribbean; introduced by Alex Tuzhilin)

Grand Challenges on the Road to Practical Data Mining Systems

Usama Fayyad, DMX Group, USA

The past two decades have seen a huge wave of computational systems for the "digitization" of business operations from ERP, to manufacturing, to systems for customer interactions. These systems increased the throughput and efficiency of conducting "transactions" and resulted in an unprecedented build-up of data captured from these systems. The paradoxical reality that most organizations face today is that they have more data about every aspect of their operations and customers, yet they find themselves with an ever diminishing understanding of either. Data Mining has received much attention as a technology that can possibly bridge the gap between data and knowledge. While some interesting progress has been achieved over the past few years, especially when it comes to techniques and scalable algorithms, very few organizations have managed to benefit from the technology. Despite the recent advances, some major hurdles exist on the road to the needed evolution. Furthermore, most technical research work does not appear to be directed at these challenges, nor does it appear to be aware of their nature. This talk will cover these challenges and present them in both the technical and the business context. The exposition will cover deep technical research questions, practical application considerations, and social/economic considerations. The talk will draw on illustrative examples from scientific data analysis, commercial applications of data mining in understanding customer interaction data, and considerations of coupling data mining technology within database management systems. Of particular interest are the business challenges of how to make the technology really work in practice. There are many unsolved deep technical research problems in this field and we conclude by covering a sampling of these.

9:40-10:10am Coffee Break

10:10-11:40am Technical Sessions 4

Barbados Room – Association Rules 1 (Session 4a Chair: Jian Pei)

10:10am (22 min)	*Mining strong affinity association patterns in data sets with skewed support distribution,* by Hui Xiong, Pang-Ning Tan, and Vipin Kumar
10:32am (11 min)	*Objective and subjective algorithms for grouping association rules*, by Aijun An, Shakil Khan, and Xiangji Huang
10:43am (11 min)	*Interpretations of association rules by granular computing,* by Yuefeng Li and Ning Zhong
10:54am (22 min)	*A high-performance distributed algorithm for mining association rules,* by Assaf Schuster, Ran Wolff, and Dan Trock
11:16am (22 min)	*Association rule mining in peer-to-peer systems,* by Ran Wolff and Assaf Schuster

Aruba Room – Text 2 (Session 4b Chair: Larry Hall)

10:10am (22 min)	*Sentiment analyzer: Extracting of sentiments towards a given topic using natural language processing techniques,* by Jeonghee Yi, Tetsuya Nasukawa, R. Bunescu, and W. Niblack
10:32am (22 min)	*CBC: Clustering based text classification requiring minimal labeled data,* by H.-J. Zeng, X.-H. Wang, Z. Chen, H. Lu and W.-Y. Ma
10:54am (22 min)	*Parsing without a grammar: Making sense of unknown file formats,* by Levon Lloyd and Steven Skiena

Trinidad Room – Data Cleaning & Methodological Issues (Session 4c Chair: Bing Liu)

10:10am (22 min)	*Is random model better? On its accuracy and efficiency,* by Wei Fan, Haixun Wang, Philip Yu, and Sheng Ma
10:32am (11 min)	*Impact studies and sensitivity analysis in medical data mining with ROC-based genetic learning,* by Michele Sebag, Jerome Azé, and N. Lucas
10:43am (11 min)	*Comparing naive Bayes, decision trees, and SVM with AUC and Accuracy*, by Jin Huang, J. Lu, and Charles Ling
10:54am (22 min)	*Probabilistic noise identification and data cleaning,* by Jeremy Kubica and Andrew Moore
11:16am (11 min)	*Algorithms for spatial outlier detection,* by Chang-Tien Lu, Dechang Chen, and Yufeng Kou
11:27am (11 min)	*Model stability: A key factor in determining whether an algorithm produces an optimal model from a matching distribution,* by Kai Ming Ting and R. J. Y. Quek

Penthouse – Feature Selection (Session 4d Chair: Hiroshi Motoda)

10:10am (22 min)	*Efficient nonlinear dimension reduction for clustered data using kernel functions,* by Cheong Hee Park and Haesun Park
10:32am (11 min)	*Active sampling for feature selection,* by Sriharsha Veeramachaneni and Paolo Avesani
10:43am (11 min)	*A feature selection framework for text filtering,* by Zhaohui Zheng, Rohini Srihari, and Sargur Srihari
10:54am (11 min)	*Analyzing high-dimensional data by subspace validity,* by Amihood Amir, Reuven Kashi, Nathan Netanyahu, Daniel Keim, and Markus Wawryniuk
11:05am (11 min)	*A fast algorithm for computing hypergraph transversals and its application in mining emerging patterns,* by James Bailey, Thomas Manoukian, and Kotagiri Ramamohanarao
11:16am (11 min)	*Dimensionality reduction using kernel pooled local discriminant information*, by Peng Zhang, Jing Peng, and Carlotta Domeniconi

11:40-12:00pm Box Lunch on Oceanfront Deck (included in registration)

12:00-6:30pm Tour of NASA Kennedy Spaceflight Center

6:30-10:00pm Banquet in Grand Caribbean Ballroom

ICDM 2003 Saturday (November 22, 2003)

8:30-9:40am Invited Talk (Grand Caribbean; introduced by Xindong Wu)

Sequential Supervised Learning:
General Methods for Sequence Labeling and Segmentation

Thomas Dietterich, Oregon State University, USA

Many existing and emerging applications of machine learning and data mining involve the problem of labeling the elements of a sequence. Examples include information extraction from web pages, part-of-speech tagging in computational linguistics, protein and DNA sequence analysis, and computer intrusion detection. In all of these tasks, the training examples consist of pairs (X,Y), where X is a sequence of objects or events $(x_1, ..., x_T)$ each described by a vector of features, and Y is a matching sequence of class labels $(y_1, ..., y_T)$. Given a new sequence of objects X, the goal is to predict the corresponding sequence of labels Y. This is an example of "collective classification", where each of the objects x_t is classified simultaneously with all of the other objects in the sequence. This talk will discuss practical, off-the-shelf machine learning methods for sequential supervised learning and describe our experience with applications in computational linguistics, information extraction, and bioinformatics.

9:40-10:10am Coffee Break

10:10-12:00pm Technical Sessions 5

Barbados Room – Part A: Privacy-Preserving Data Mining,
Part B: Databases and Data Mining (Session 5a Chair: Chris Clifton)

10:10am (22 min)	*On the privacy preserving properties of random data perturbation techniques*, by Hillol Kargupta, Souptik Datta, Qi Wang, and Krishnamoorthy Sivakumar
10:32am (11 min)	*Protecting sensitive knowledge by data sanitization*, by Stanley Oliveira and Osmar Zaïane
10:43am (22 min)	*Privacy-preserving distributed clustering using generative models*, by Srujana Merugu and Joydeep Ghosh
11:05am (11 min)	*Privacy-preserving collaborative filtering using randomized perturbation techniques*, by Huseyin Polat and Wenliang Du
11:16am (22 min)	*An algebra for inductive query evaluation*, by Sau Dan Lee and Luc De Raedt
11:38am (22 min)	*Zigzag: A new algorithm for mining large inclusion dependencies in databases*, by Fabien De Marchi and Jean-Marc Petit

Aruba Room – Mining Sequential and Hierarchical Data
(Session 5b Chair: Michele Sebag)

10:10am (22 min) *Reliable detection of episodes in event sequences*, by Robert Gwadera, Mikhail Atallah, and Wojciech Szpankowski

10:32am (11 min) *Enhancing techniques for efficient topic hierarchy integration*, by Jyh-Jong Tsay, C.-F. Chang, H.-Y. Chen, and C.-H. Lin

10:43am (22 min) *Sequence modeling with mixtures of conditional maximum entropy distributions*, by Dmitry Pavlov

11:05am (22 min) *TSP: Mining top*-k *closed sequential patterns*, by Petre Tzvetkov, Xifeng Yan, and Jiawei Han

11:27pm (11 min) *Mining semantic networks for knowledge discovery*, by Kanagasabai Rajaraman and Ah-Hwee Tan

11:38am (22 min) *Introducing uncertainty into pattern discovery in temporal event sequences*, by Xingzhi Sun, Maria E. Orlowska, and Xue Li

Trinidad Room – Mining Frequent Items (Session 5c Chair: Qiang Yang)

10:10am (22 min) *ExAMiner: Optimized level-wise frequent pattern mining with monotone constraints*, by Francesco Bonchi, Fosca Giannotti, Alessio Mazzanti, and Dino Pedreschi

10:32am (22 min) *Frequent sub-structure-based approaches for classifying chemical compounds*, by Mukund Deshpande, Michihiro Kuramochi, and George Karypis

10:54am (11 min) *Efficient Mining of frequent subgraph in the presence of isomorphism*, by Jun Huan, Wei Wang, and Jan Prins

11:05am (22 min) *Efficient data mining for maximal frequent subtrees*, by Yongqiao Xiao, Jenq-Foung Yao, Zhigang Li, and Margaret Dunham

11:27am (22 min) *Mining high utility itemsets*, by Raymond Chan, Qiang Yang, and Yi-Dong Shen

11:49am (11 min) *Mining frequent itemsets in distributed and dynamic databases*, by Matthew Otey, Chao Wang, Srinivasan Parthasarathy, Adriano Veloso, and Wagner Meira Jr.

Penthouse – Clustering 3 (Session 5d Chair: Guozhu Dong)

10:10am (22 min) *MaPle: A fast algorithm for maximal pattern-based clustering*, by Jian Pei, Xiaoling Zhang, Moonjung Cho, Haixun Wang, and Philip Yu

10:32am (22 min) *Scalable model-based clustering by working on data summaries*, by Huidong Jin, Man-Leung Wong, and Kwong-Sak Leung

10:54am (22 min) *Combining multiple weak clusterings*, by Alexander Topchy, Anil Jain, and William Punch

11:16am (22 min) *Regression clustering*, by Bin Zhang

11:38am (22 min) *TECNO-STREAMS: Tracking evolving clusters in noisy data streams with a scalable immune system learning model,* by Olfa Nasraoui, Cesar Cardona Uribe, Carlos Rojas Coronel, and Fabio Gonzalez

12:00-1:30pm Lunch (on Oceanfront Deck; included in registration)
12:45-1:30pm TCCI Business Meeting (Grand Caribbean Ballroom)

1:30-2:40pm Invited Talk (Grand Caribbean; introduced by Ning Zhong)

Global Structure from Sequences

Heikki Mannila, University of Helsinki, Finland

Sequences of discrete symbols or continuous values occur in many applications, such as bioinformatics and process monitoring. In this talk we describe some ways of finding global structure from sequences. First, we consider finding recurrent sources in sequences, i.e., identifying h possible sources such that the sequence can be written as a concatenation of $k>h$ pieces, each of which stems from one of the h sources. Second, we describe some approaches to modeling the intensities of events in sequences of events using dynamic programming and reversible jump Markov chain Monte Carlo. Both methods are applied to biological sequences.
(Joint work with Aris Gionis and Marko Salmenkivi.)

2:45-3:45pm Panel and Technical Session 6

Barbados Room – Panel on Security and Data Mining

This panel will discuss the opportunities for funding available for data mining research and applications. Panelists will discuss the priorities within their organization and how data mining research fits into the larger context of the organization. Applications and challenges for security applications will be discussed as well as applications that enable advances in science.

Panelists: Michael Pazzani (panel chair) of United States National Science Foundation (NSF), Kei Koizumi of the American Association for the Advancement of Science (AAAS), and Rick Steinheiser, United States Central Intelligence Agency (CIA).

Trinidad Room – Support Vector Machines and Nearest-Neighbor Methods

(Session 6 Chair: Hillol Kargupta)

2:45pm (11 min)	*SVM based models for predicting foreign currency exchange rates*, by Joarder Kamruzzaman, Ruhul Sarker, and I. Ahmad
2:56pm (11 min)	*An algorithm for the exact computation of the centroid of higher dimensional polyhedra and its application to kernel machines*, by Frederic Maire
3:07pm (11 min)	*K-D decision tree: An accelerated and memory efficient nearest neighbor classifier*, by Tomoyuki Shubata, Takekazu Kato, and Toshikazu Wada
3:18pm (11 min)	*Center-based indexing for nearest neighbors search*, by Arkadiusz Wojna
3:29pm (11 min)	*A K-NN associated fuzzy evidential reasoning classifier with adaptive neighbor selection*, by Hongwei Zhu and Otman Basir

3:45-4:15pm Coffee Break

4:15-5:45pm Technical Sessions 7

Barbados Room – Linkage-Based Methods (Session 7a Chair: Mark Maloof)

4:15pm (22 min)	*Unsupervised link discovery in multi-relational data via rarity analysis*, by Shou-de Lin and Hans Chalupsky
4:37pm (11 min)	*Tractable group detection on large link data sets*, by Jeremy Kubica, Andrew Moore, and Jeff Schneider
4:48pm (22 min)	*Mining significant pairs of patterns from graph structures with class labels*, by Akihiro Inokuchi and Hisashi Kashima
5:10pm (11 min)	*Links between Kleinbergs hubs and authorities, correspondence analysis, and Markov chains*, by Francois Fouss, Marco Saerens, and Jean-Michel Renders
5:21pm (22 min)	*Identifying Markov blankets with decision tree induction*, by Lewis Frey, Douglas Fisher, Ioannis Tsamardinos, Constantin Aliferis, and Alexander Statnikov

Aruba Room – Part A: Applications Track 2, Part B: Bioinformatics
(Session 7b Chair: Chris Ding)

4:15pm (11 min)	*Mining production data with neural network and CART*, by Mingkun Li, Shuo Feng, Ishwar Sethi, Jason Luciow, and Keith Wagner
4:26pm (11 min)	*Findings from a practical project concerning web usage mining*, by Frank Dellmann, Holger Wulff, and Stefan Schmitz
4:37pm (11 min)	*Text mining for a clear picture of defect reports: A praxis report*, by Jutta Kreyss, Steve Selvaggio, Michael White, and Zach Zakharian
4:48pm (11 min)	*Detecting patterns of change using enhanced parallel coordinate visualization*, by Kaidi Zhao, Bing Liu, Tom Tirpak, and Andreas Schaller
4:59pm (11 min)	*A hybrid data-mining approach in genomics and text structures*, by Horia Nicolai Teodorescu and LucianIulian Fira
5:10pm (11 min)	*Effectiveness of information extraction, multi-relational, and semi-supervised learning for predicting functional properties of genes*, by Mark Krogel and Tobias Scheffer

Trinidad Room – Association Rules 2 (Session 7c Chair: Chang-Tien Lu)

4:15pm (22 min)	*Optimized disjunctive association rules via sampling*, by Joseph Elble, Cinda Heeren, and Leonard Pitt
4:37pm (11 min)	*Integrating fuzziness into OLAP for multidimensional fuzzy association rules mining*, by Reda Alhajj and Mehmet Kaya
4:48pm (22 min)	*Change profiles*, by Taneli Mielikdinen
5:10pm (11 min)	*CoMine: Efficient Mining of Correlated Patterns*, by Young-Koo Lee, Won-Young Kim, Y. Dora Cai, and Jiawei Han
5:21pm (11 min)	*The Rough Set Approach to Association Rule Mining*, by Jiwen Guan, David Bell, and D. Y. Liu

Penthouse – Visualization 2 and Image Processing (Session 7d Chair: Ian Davidson)

4:15pm (22 min)	*Detecting interesting exceptions from medical test data with visual summarization*, by Einoshin Suzuki, Takeshi Watanabe, Hideto Yokoi, and Katsuhiko Takabayashi
4:37pm (22 min)	*Spatial Interest Pixels (SIPs): Useful low-level features of visual media data*, by Qi Li, Jieping Ye, and Chandra Kambhamettu
4:59pm (22 min)	*Evolutionary Gabor filter optimization with application to vehicle detection*, by Zehang Sun, George Bebis, and Ronald Miller

The Hotel's Floor Plan