Program for the

Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

KDD 2002

Edmonton, Alberta, Canada
July 23-26, 2002

Program Highlights

Invited Talks

- Keynote address by Daryl Pregibon, AT&T Shannon labs: “Where to from here?”.
- Keynote speech by Geoffrey Hinton, Department of Computer Science, University of Toronto: “Data-Mining The Optic Nerve”.

Research and Industrial Tracks

- 32 research papers divided into two tracks with thirteen sessions in total.
- 12 industrial papers divided in four sessions.
- 44 poster papers
- one panel session: “The perfect data mining tool: Automated or Interactive ?”
- 6 tutorials:
    - Multivariate Density Estimation and Visual Clustering.
    - Text Mining for Bioinformatics.
    - Link Analysis : Current State of the Art.
    - Common Reasons Data Mining Projects Fail.
    - Querying and Mining Data Streams: you only get one look.
    - Visual Data Mining: Background, Techniques, and Drug Discovery Applications.
- 6 Workshops:
    - 4th WEBKDD: Web Mining for Usage Patterns and User Profiles.
    - 3rd MDM/KDD: Workshop on Multimedia Data Mining.
    - 2nd BIOKDD: Workshop on Data Mining in Bioinformatics.
    - 2nd Workshop on Temporal Data Mining.
    - MRDM Multi-Relational Data Mining.
    - Fractals and Self-similarity in Data Mining: Issues and Approaches.

Tuesday, July 23

8:00-17:00

Registration (Turner Valley)

9:00-10:00

Tutorial: Text Mining for Bioinformatics (Manitoba)

2nd BIOKDD: Workshop on Data Mining in Bioinformatics (British Columbia)

4th WEBKDD: Web Mining for Usage Patterns and User Profiles (Alberta)

3rd MDM/KDD: Workshop on Multimedia Data Mining (Yukon)

MRDM Multi-Relational Data Mining (Strathcona)

2nd Workshop on Temporal Data Mining (Chairman)

10:00-10:30

Coffee Break (N/W Foyer)

10:30-12:00

Tutorial: Text Mining for Bioinformatics (Manitoba)

2nd BIOKDD: Workshop on Data Mining in Bioinformatics (British Columbia)

4th WEBKDD: Web Mining for Usage Patterns and User Profiles (Alberta)

3rd MDM/KDD: Workshop on Multimedia Data Mining (Yukon)

MRDM Multi-Relational Data Mining (Strathcona)

2nd Workshop on Temporal Data Mining (Chairman)

12:00-13:30

Break

Tuesday, July 23

13:30-16:00

Tutorial: Link Analysis: Current State of the Art (Manitoba)

Tutorial: Querying and Mining Data Streams: you only get one look (Saskatchewan)

2nd BIOKDD: Workshop on Data Mining in Bioinformatics (British Columbia)

4th WEBKDD: Web Mining for Usage Patterns and User Profiles (Alberta)

3rd MDM/KDD: Workshop on Multimedia Data Mining (Yukon)

MRDM Multi-Relational Data Mining (Strathcona)

2nd Workshop on Temporal Data Mining (Chairman)

Workshop on Fractals and Self-similarity in Data Mining: Issues and Approaches (Leduc)

16:00-17:00

Coffee Break (N/W Foyer)

17:00-17:15

KDD 2002 Opening (Ballroom)

17:15-18:15

Best Paper Presentations (Ballroom)
Chair: Hans-Peter Kriegel

Pattern Discovery in Sequences under a Markov Assumption
(Best Research Paper)

Darya Chudova, Padhraic Smyth

Customer Lifetime Value Modeling and Its Use for Customer Retention Planning
(Best Application Paper)
Saharon Rosset, Einat Neumann, Uri Eick, Nurit Vatnik, Yizhak Idan

18:15-19:15

KDD Cup Awards (Ballroom)
Chairs: Mark Craven, Alexander Yeh

Wednesday, July 24

7:30-8:30

Continental Breakfast (N/W Foyer)

8:00-17:00

Registration (Turner Valley)

8:00-17:00

Exhibits (N/W Foyer)

8:30-10:00

Keynote Talk (Ballroom)
Chair: David Hand

Where to from here? - Daryl Pregibon, AT&T Shannon labs

The emergence and success of KDD and data mining is the result of gaps between the fields of machine learning, statistics, database research, and artificial intelligence. Like KDD, all of these fields have evolved over the past decade, and it is reasonable to assess the distinctiveness of KDD to its more mature siblings. We argue that since change is inevitable, we must control what we can, and influence what we can't. To do otherwise leaves the destiny of KDD to other forces, whether they are technological, societal, or political.
We discuss some recent trends and events, e.g., the dot com meltdown, and some ways for the field to respond to the challenges, and the opportunities.

10:00-10:30

Coffee Break (N/W Foyer)

Wednesday, July 24

10:30-12:00 Research Track 1

Statistical Methods I (British Columbia/Alberta/Yukon)
Chair: Eamonn Keogh

Bayesian analysis of massive datasets via particle filters Greg Ridgeway, David Madigan	Scalable Robust Covariance and Correlation Estimates for Data Mining Fatemah A. ALqallaf, Kjell P. Konis, R. Douglas Martin, Ruben H. Zamar	MARK: A Boosting Algorithm for Heterogeneous Kernel Models Kristin P. Bennett, Michinari Momma, Mark J. Embrechts

10:30-12:00 Research Track 2

Frequent Patterns I (Manitoba)
Chair: David Jensen

Selecting the Right Interestingness Measure for Association Patterns Pang-Ning Tan, Vipin Kumar, Jaideep Srivastava	DualMiner: A Dual-Pruning Algorithm for Itemsets with Constraints Cristian Bucila, Johannes Gehrke, Daniel Kifer, Walker White	Querying Multiple Sets of Discovered Rules Alexander Tuzhilin, Bing Liu

10:30-12:00 Research Track 3

Graphs and Trees (Saskatchewan)
Chair: Minos Garofalakis

Mining Knowledge-Sharing Sites for Viral Marketing Matthew Richardson, Pedro Domingos	Efficiently Mining Frequent Trees in a Forest Mohammed Zaki	ANF: A Fast and Scalable Tool for Data Mining in Massive Graphs Christopher R. Palmer, Phillip B. Gibbons, Christos Faloutsos

12:00-13:30

Lunch (N/W Foyer)

Wednesday, July 24

13:30-14:30 Research Track 1

Streams and Time Series
(British Columbia/Alberta/Yukon) **Chair:** Greg Ridgeway

Bursty and Hierarchical Structure in Streams

Jon Kleinberg

On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration

Eamonn Keogh, Shruti Kasetty

13:30-14:30 Research Track 2

Visualization (Manitoba)
Chair: Georges Grinstein

Query, Analysis, and Visualization of Hierarchically Structured Data using Polaris

Christopher Stolte, Diane Tang, Pat Hanrahan

On Interactive Visualization of high-dimensional Data using the Hyperbolic Plane

Jörg Walter, Helge Britter

13:30-14:30 Tutorial

Tutorial: Common Reasons Data Mining Projects Fail (Saskatchewan)

14:30-15:00

Coffee Break (N/W Foyer)

Wednesday, July 24

15:00-16:00 Research Track 1

Web Search and Navigation
(British Columbia/Alberta/Yukon) **Chair**: Alex Tuzhilin

Optimizing Search Engines Using Clickthrough Data

Thorsten Joachims

Relational Markov Models and their Application to Adaptive Web Navigation

Corin R. Anderson, Pedro Domingos, Daniel S. Weld

15:00-16:00 Research Track 2

Ensembles and Boosting (Manitoba)
Chair: Simeon Simoff

Predicting Rare Classes: Can Boosting Make Any Weak Learner Strong?

Mahesh V. Joshi, Ramesh C. Agarwal, Vipin Kumar

Efficient Handling of High-Dimensional Feature Spaces by Randomized Classifier Ensembles

Aleksander Kolcz, Xiaomei Sun, Jugal Kalita

15:00-16:30 Tutorial

Tutorial: Common Reasons Data Mining Projects Fail (Saskatchewan)

16:30-18:00

Plenary Poster Presentations (Ballroom)
Chair: Raghu Ramakrishnan

44 Posters

18:00-20:00

Poster Session and Reception (Pradera Café)

Poster Papers

Collaborative Crawling: Mining User Experiences for Topical Resource Discovery
Charu C. Aggarwal

Sequential PAttern Mining Using Bitmap Representation
Jay Ayres, Jason Flannick, Johannes Gehrke, Tomi Yiu

Frequent Term-Based Text Clustering
Florian Beil, Martin Ester, Xiaowei Xu

A Theoretical Framework for Learning from a Pool of Disparate Data Sources
Shai Ben-David, Johannes Gehrke, Reba Schuller

Topics in 0-1 data
Ella Bingham, Heikki Mannila, Jouni K. Seppanen

Extracting Decision Trees From Trained Neural Networks
Olcay Boz

A New Two-Phase Sampling Based Algorithm for Discovering Association Rules
Bin Chen, Peter Haas, Peter Scheuermann

CVS: A Correlation-Verification Based Smoothing Technique on Information Retrieval and Term Clustering
Christina Yip Chung, Bin Chen

Learning to Match and Cluster Large High-Dimensional Data Sets For Data Integration
William W. Cohen, Jacob Richman

SECRET: A Scalable Linear Regression Tree Algorithm
Alin Dobra, Johannes Gehrke

Statistical Modeling of Large-Scale Simulation Data
Tina Eliassi-Rad, Terence Critchlow, Ghaleb Abdulla

Tumor Cell Identification using Features Rules
Bin Fang, Wynne Hsu, Mong Li Lee

Integrating Feature and Instance Selection for Text Classification
Dimitris Fragoudis, Dimitris Meretakis, Spiros Likothanassis

SyMP: An Efficient Clustering Approach to Identify Clusters of Arbitrary Shapes in Large Data Sets
Hichem Frigui

Poster Papers (cont)

Scaling multi-class Support Vector Machines using inter-class confusion
Shantanu Godbole, Sunita Sarawagi, Soumen Chakrabarti

Visualization Support for an User-Centered KDD Process
TuBao Ho, TrongDung Nguyen, DungDuc Nguyen

Mining Complex Models from Arbitrarily Large Databases in Constant Time
Geoff Hulten, Pedro Domingos

A Model for Discovering Customer Value for E-Content
Srinivasan Jagannathan, Jayanth Nayak, Kevin Almeroth, Markus Hofmann

SimRank: A Measure of Structural-Context Similarity
Glen Jeh, Jennifer Widom

Similarity Measure Based on Partial Information of Time series
Xiaoming Jin, Yuchang Lu, Chunyi Shi

Finding Surprising Patterns in a Time Series Database In Linear Time and Space
Eamonn Keogh, Stefano Lonardi, Bill Yuan-chi Chiu

Clustering Seasonality Patterns in the Presence of Errors
Mahesh Kumar, Nitin R. Patel, Jonathan Woo

Construct robust rule sets for classification
Jiuyong Li, Rodney Topor, Hong Shen

Instability of Decision Tree Classification Algorithms
Ruey-Hsia Li, Geneva G. Belford

Distributed Data Mining in a Chain Store Database of Short Transactions
Cheng-Ru Lin, Chang-Hung Lee, Ming-Syan Chen, Philip S. Yu

A Robust and Efficient Clustering Algorithm based on Cohesion Self-Merging
Cheng-Ru Lin, Ming-Syan Chen

Discovering Informative Content Blocks from Web Documents
Shian-Hua Lin, Jan-Ming Ho

Poster Papers (cont)

Collusion in The U.S. Crop Insurance Program: Applied Data Mining
Bert B. Little, Walter L. Johnston, Ashley C. Lovell, Roderick M. Rejesus, Steve A. Steed

Incremental Context Mining for Adaptive Document Classification
Rey-Long Liu, Yun-Ling Lu

Evaluating Classifiers' Performance in a Constrained Environment
Anna Olecka

Discovering Word Senses from Text
Patrick Pantel, Dekang Lin

Combining Clustering and Co-training to Enhance Text Classification Using Unlabelled Data
Bhavani Raskutti, Herman Ferra, Adam Kowalczyk,

Single-shot Detection of Multiple Categories of Text using Parametric Mixture Models
Naonori Ueda, Kazumi Saito

What's the Code? Automatic Classification of Source Code Archives
Secil Ugurel, Robert Krovetz, Lee Giles, David M. Pennock, Eric J. Glover, Hongyuan Zha

Privacy Preserving Association Rule Mining in Vertically Partitioned Data
Jaideep Vaidya, Chris Clifton

Non-Linear Dimensionality Reduction Techniques for Classification and Visualization
Michail Vlachos, Carlotta Domeniconi, Dimitris Gunopulos, George Kollios, Nick Koudas

Item Selection By "Hub-Authority" Profit Ranking
Ke Wang, Ming-Yen Thomas Su

Discovery Net: Towards a Grid of Knowledge Discovery
Vasa Curcin, Moustafa Ghanem, Yike Guo, Martin Kohler,Anthony Rowe, Jameel Syed, Patrick Wendel

Making every bit count: Fast nonlinear axis scaling
Leejay Wu, Christos Faloutsos

B-EM: A Classifier Incorporating Bootstrap with EM Approach for Data Mining
Xintao Wu, Jianping Fan, Kalpathi R. Subramanian

Poster Papers (cont)

A Unifying Framework for Detecting Outliers and Change Points from Non-stationary Time Series Data
Kenji Yamanishi, Jun-ichi Takeuchi

CLOPE: A Fast and Effective Clustering Algorithm for Transactional Data
Yiling Yang, Xudong Guan, Jinyuan You

Topic-conditioned Novelty Detection
Yiming Yang, Jian Zhang, Jaime Carbonell, Chun Jin

Transforming classifier scores into accurate multiclass probability estimates
Bianca Zadrozny, Charles Elkan

Thursday, July 25

7:30-8:30

Continental Breakfast (N/W Foyer)

8:00-17:00

Registration (Turner Valley)

8:00-17:00

Exhibits (N/W Foyer)

8:30-10:00

Panel: The Perfect Data Mining Tool: Interactive or Automated? (Manitoba)
Chair: Mihael Ankerst, The Boeing Company

Panelists:

- Surajit Chaudhuri, Microsoft Research
- Georges Grinstein, University of Massachusetts Lowell & AnVil, Inc.
- Jiawei Han, University of Illinois at Urbana-Champaign
- Gregory Piatetsky-Shapiro, KDnuggets

Commercial and academic data mining tools range from being fully automated to highly interactive. We will discuss the role of human involvement in the data mining process. On the one hand, providing interactivity/visualization enables domain knowledge transfer and the use of the human-perceptual capabilities. On the other hand, the vast amount of data to be mined today makes real-time interactivity hard to achieve and unnecessarily burdens the user to perform tasks that may be done automatically. Questions to be discussed include: What are the ideal roles of the computer and of the user in the data mining process? Which data mining methods, clustering, classification, association rules, can be improved by more human involvement? What kind of application requires more human involvement? What kind of application requires little or no human involvement?

8:30-10:00

Tutorial: Multivariate Density Estimation and Visual Clustering (British Columbia/Alberta/Yukon)

Thursday, July 25

8:30-10:00 Industrial Track

Market Intelligence and Software Mining (Saskatchewan) **Chair:** Christos Faloutsos

Learning Domain-Independent String Transformation Weights for High Accuracy Object Identification	A System for Real-time Competitive Market Intelligence	From Run-time Behavior to Usage Scenarios: An Interaction-pattern Mining Approach
Sheila Tejada, Craig A. Knoblock, Steven Minton	Sholom M. Weiss, Naval K. Verma	Mohammad El-Ramly, Eleni Stroulia, Paul Sorenson

10:00-10:30

Coffee Break (N/W Foyer)

10:30-11:30

Tutorial: Multivariate Density Estimation and Visual Clustering (British Columbia/Alberta/Yukon)

10:30-11:30 Research Track

Statistical Methods II (Manitoba)
Chair: Doug Martin

Shrinkage Estimator Generalizations of Proximal Support Vector Machines	Hierarchical Model-Based Clustering of Large Datasets Through Fractionation and Refractionation
Deepak K. Agarwal	Jeremy Tantrum, Alejandro Murua, Werner Stuetzle

10:30-11:30 Industrial Track

Customer and Product Analysis (Saskatchewan)
Chair: Greg Piatetsky-Shapiro

Exploiting Response Models - Optimizing Cross-Sell and Up-Sell Opportunities in Banking	Mining Product Reputations on the Web
Andrew Storey, Marc-david Cohen	Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, Toshikazu Fukushima

Thursday, July 25

11:45-13:45

SIGKDD Business Lunch (Meeting/Luncheon at Shaw Center)

14:00-15:30 Research Track

Text Classification (British Columbia/Alberta/Yukon)
Chair: Osmar R. Zaïane

Enhanced Word Clustering for Hierarchical Text Classification Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar	A Parallel Learning Algorithm for Text Classification Canasai Kuengkrai, Chuleerat Jaruskulchai	A Refinement Approach to Handling Model Misfit in Text Categorization Haoran Wu, Tong Heng Phang, Bing Liu, Xiaoli Li

14:00-15:30 Industrial Track

Intrusion Detection (Manitoba)
Chair: Johannes Gehrke

Mining Intrusion Detection Alarms for Actionable Knowledge Klaus Julisch, Marc Dacier	Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks Matthew V. Mahoney, Philip K. Chan	ADMIT: Anomaly-based Data Mining for Intrusions Karlton Sequeira, Mohammed Zaki

14:00-15:30

Tutorial: Visual Data Mining: Background, Techniques, and Drug Discovery Applications (Saskatchewan)

15:30-16:00

Coffee Break (N/W Foyer)

Thursday, July 25

16:00-17:30 Research Track

Web Page and String Classification
(British Columbia/Alberta/Yukon) **Chair:** Bing Liu

PEBL: Positive Example Based Learning for Web Page Classification Using SVM Hwanjo Yu, Jiawei Han, Kevin Chen-Chuan Chang	Web Site Mining: A new way to spot Competitors, Customers and Suppliers in the World Wide Web Martin Ester, Hans-Peter Kriegel, Matthias Schubert	On Effective Classification of Strings with Wavelets Charu C. Aggarwal

16:00-17:30 Industrial Track

Bioinformatics (Manitoba)
Chair: Michael Berthold

Handling Very Large Numbers of Association Rules in the Analysis of Microarray Data Alexander Tuzhilin, Gediminas Adomavicius	On the potential of domain literature for clustering and for Bayesian network learning Peter Antal, Patrick Glenisson, Geert Fannes	Mining Heterogeneous Gene Expression with Time Lagged Recurrent Neural Networks Yulan Liang, Arpad Kelemen

16:00-17:30

Tutorial: Visual Data Mining: Background, Techniques, and Drug Discovery Applications (Saskatchewan)

Friday, July 26

8:00-9:00

Continental Breakfast (N/W Foyer)

9:00-10:30

Capstone Talk (Ballroom)
Chair: Raymond Ng

Data-Mining The Optic Nerve - Geoffrey Hinton, Department of Computer Science, University of Toronto

The optic nerve supplies the brain with sensory data at millions of bits per second. This data is transformed by the early visual system into activations of a very large number of feature detectors. I will describe a multilayer, unsupervised learning algorithm, developed in collaboration with Max Welling, Yee-Whye Teh and Simon Osindero, that discovers a set of feature detectors that form a good probability density model of the sensory input. These feature detectors look remarkably like those that exist in the brain. The learning algorithm works by trying to find many different soft constraints in the data. When a constraint is satisfied it does not have to be satisfied exactly and when it is violated it can be violated by a very large amount. The algorithm can be viewed as a generalization of independent component analysis and also as a generalization of maximum entropy methods that have been used for modelling word strings.

10:30-11:00

Coffee Break (N/W Foyer)

11:00-12:00 Research Track 1

Frequent Patterns II (British Columbia/Alberta/Yukon)
Chair: Chabane Djeraba

Privacy Preserving Mining of Association Rules

Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, Johannes Gehrke

Mining Frequent Item Sets by Opportunistic Projection

Junqiang Liu, Yunhe Pan, Ke Wang, Jiawei Han

Friday, July 26

11:00-12:00 Research Track 2

Learning Methods (Manitoba)
Chair: Mihael Ankerst

Sequential Cost-Sensitive Decision Making with Reinforcement Learning

Edwin Pednault, Naoki Abe, Bianca Zadrozny

Interactive Deduplication using Active Learning

Sunita Sarawagi, Anuradha Bhamidipaty

11:00-12:00 Research Track 3

Intrusion and Privacy (Saskatchewan)
Chair: Jörg Sander

Transforming Data to Satisfy Privacy Constraints

Vijay S. Iyengar

Exploiting Unlabeled Data in Ensemble Methods

Kristin P. Bennett, Ayhan Demiriz, Richard

SIGKDD Conference Organizing Committee

General Chair

Randy Goebel, University of Alberta, Canada

Program Co-Chairs

David Hand, Imperial College, UK
Daniel Keim, AT&T, USA
Raymond Ng, University of British Columbia, Canada

Industrial Session Co-Chairs

Michael Berthold, Tripos, USA
William DuMouchel, AT&T, USA

Proceedings Chair

Osmar R. Zaïane, University of Alberta, Canada

Best Paper Award Chair

Hans-Peter Kriegel, University of Munich, Germany

Panels Chair

Rajeev Rastogi, Bell Labs, USA

Treasurer

Davood Rafiei, University of Alberta, Canada

Local Arrangements Chair

Mario A. Nascimento, University of Alberta, Canada

Student Awards Chair

Jiawei Han, University of Illinois, USA

Publicity Chair

Osmar R. Zaïane, University of Alberta, Canada

Exhibits Chair

Nick Koudas, AT&T, USA

Workshops Chair

Renée J. Miller, University of Toronto, Canada

SIGKDD Conference Organizing Committee

Tutorials Chair

Alexander Hinneburg, University of Halle, Germany

Registration Chair

Jörg Sander, University of Alberta, Canada

KDD Cup Co-Chairs

Mark Craven, University of Wisconsin, USA
Alexander Yeh, The MITRE Corporation, USA

Sponsorship Co-Chairs

Charu Aggarwal, IBM Watson, USA
Nick Cercone, University of Waterloo, Canada
Mario Schkolnik, MS3, USA

Web-Site Chair

Osmar R. Zaïane, University of Alberta, Canada

SIGKDD Program Committee

Niall Adams, Imperial College, UK
Jesús S. Aguilar-Ruiz, University of Seville, Spain
Mihael Ankerst, Boeing, USA
Daniel Barbara, George Mason University, USA
Roberto Bayardo, IBM Almaden, USA
Michael Berthold, Tripos, Ca, USA
Richard Bolton, Imperial College, UK
Paul Bradley, DigiMine, USA
Nick Cercone, University of Waterloo, Canada
David Cheung, HKU, Hong Kong
Ken Church, AT&T, USA
Di Cook, Iowa State University, USA
Gautam Das, Microsoft, USA
Umeshwar Dayal, HP Labs, USA
Victor DeGruttola, Harvard School of Public Health, USA
Vasant Dhar, NYU, USA
Chabane Djeraba, University of Nantes, France
Maggie Dunham, Southern Methodist University, USA
Martin Ester, Simon Fraser University, Canada
Christos Faloutsos, CMU, USA
Tom Fawcett, Hewlett Packard, USA
Usama Fayyad, DigiMine, USA
Ronen Feldman, ClearForest Corp., USA
Ada Fu, CUHK, Hong Kong
Alex Gammerman, Royal Holloway College, UK
Minos Garofalakis, Bell Labs, USA

SIGKDD Program Committee (cont)

Johannes Gehrke, Cornell U, USA
George Grinstein, University of Massachusetts, USA
Dimitris Gunopulos, UC Riverside, USA
Yike Guo, Imperial College, UK
Howard Hamilton, University of Regina, Canada
Monte Hancock, CSI Inc., USA
Joann Harvey, Schering-Plough, USA
Howard Ho, IBM Almaden Research Center, USA
Adele Howe, Colorado State University, USA
William Hsu, Kansas State University, USA
Alfred Inselberg, University of Tel Aviv, IBM Israel
David Jensen, University of Mass, USA
Ted Johnson, AT&T, USA
Laveen Kanal, University of Maryland, USA
Jon Kleinberg, Cornell University, USA
Ronny Kohavi, Blue Martini Software,USA
Flip Korn, AT&T, USA
Vipin Kumar, University of Minnesota, USA
Diane Lambert, Bell Labs, USA
Jose Latorre, University of Barcelona, Spain,
Doheon Lee, Korea Advanced Institute of Science and Technology, Korea
Bing Liu, National University Singapore
Xiao Hui Liu, Brunel University, UK
Hongjun Lu, HKUST, Hong Kong
David Madigan, AT&T, USA
Heikki Mannila, Nokia, Finland
Doug Martin, University of Washington, USA
Stan Matwin, University of Ottawa, Canada
Shinichi Morishita, University of Tokyo, Japan
Sally Morton, RAND Corporation, USA
Rajeev Motwani, Stanford U, USA
Stephen North, AT&T, USA
Doug Nychka, National Center for Atmospheric Research, USA
Gregory Piatetski-Shapiro, Xchange, USA
Daryl Pregibon, AT&T, USA
Foster Provost, NYU, USA
Prabhakar Raghavan, Verity/Stanford, USA
Raghu Ramakrishnan, Wisconsin, USA
Marco Ramoni, Harvard, USA
Greg Ridgeway, RAND Corporation, USA
Jörg Sander, University of Alberta, Canada
Matt Schonlau, RAND Corporation, USA
Dale Schuurmans, University of Waterloo, Canada
Bernhard Seeger, University of Marburg, Germany
Kyuseok Shim, Seoul National University, Korea
Roberta Siciliano, University of Naples, Italy
Simeon J. Simoff, University of Technology Sydney, Australia
Padhraic Smyth, UC, Irvine
Sal Stolfo, Columbia U, USA
Paul Stolorz, NASA, USA

SIGKDD Program Committee (cont)

Hannu Toivonen, Nokia, Finland
Alexander Tuzhilin, NYU, USA
K.P.Unnikrishnan, General Motors Research, USA
Ramasamy Uthurusamy, General Motors Corporation, USA
Matthew Ward, Worcester Polytechnic, USA
Ed Wegman, George Mason University, USA
Xindong Wu, University of Vermont, USA
Xiang Yang, University of Guelph, Canada
Hwan-Seung Yong, Ewha Womans University,Korea
Ning Zhong, Maebashi Institute of Technology, Japan

Best Paper Award Committee

Umeshwar Dayal, HP Labs, USA
Martin Ester, Simon Fraser University, Canada
Dimitrios Gunopulos, UC Riverside, USA
David Madigan, Rutgers University, USA
Heikki Mannila, University of Helsinki, Finland
Dennis Shasha, New York University, USA

ACM SIGKDD Chair

Won Kim, Cyber Database Solutions, USA

Acknowledgements

The SIGKDD 2002 Conference gratefully acknowledges the contributions of the following institutions.

Gold Sponsors:

Silver Sponsors:

Other Sponsors

Sponsoring Organizations:

Summarized Technical Program

Tuesday (Morning)

- Tutorial - Text Mining for Bioinformatics

Tuesday (Full Day)

Workshops

- 4th WEBKDD
- 3rd MDM/KDD
- 2nd BIOKDD
- 2nd Workshop on Temporal Data Mining
- MRDM

Tuesday (Afternoon)

- Tutorial - Link Analysis : Current State of the Art
- Workshop on Fractals and Self-similarity in Data Mining
- SIGKDD 2002 Opening
- Best Paper presentations (2 Papers)
- KDD Cup 2002

Wednesday (Morning)

- Plenary Session (keynote talk)
- 3 sessions in 3 Parallel Research Tracks
  - Statistical Methods I (3 papers)
  - Frequent Patterns I (3 papers)
  - Graphs and Trees I (3 papers)

Wednesday (Full Day)

- Exhibits

Wednesday (Afternoon)

- Tutorial - Common Reasons Data Mining Projects Fail
- 4 sessions in 2 Parallel Research Tracks
  - Stream and Time Series (2 papers)
  - Visualization (2 papers)
  - Web Search and Navigation (2 papers)
  - Ensembles and Boosting (2 papers)
- Poster paper presentation
- Poster session and reception

Summarized Technical Program

Thursday (Morning)

- Panel Session
- Tutorial - Multivariate Density Estimation and Visual Clustering
- 1 Research Track Session
  - Statistical Methods II (2 papers)
- 2 Industrial Track Sessions
  - Market Intelligence and Software Mining (3 papers)
  - Customer and Product Analysis (2 papers)
- SIGKDD Meeting/Luncheon at Shaw Center

Thursday (Full Day)

- Exhibits

Thursday (Afternoon)

- Tutorial - Visual DM: Background, Techniques, and Drug Discovery Applications
- 2 Research Track Sessions
  - Text Classification (3 papers)
  - Web Page and String Classification (3 papers)
- 2 Industrial Track Sessions
  - Intrusion Detection (3 papers)
  - Bioinformatics (3 papers)

Friday Morning

- Plenary Session (capstone talk)
- 3 sessions in 3 Parallel Research Tracks
  - Frequent Patterns II (2 papers)
  - Learning Methods (2 papers)
  - Intrusion and Privacy (2 papers)