The First International Conference on Knowledge Discovery and Data Mining (KDD95)

Sponsored by AAAI and in Cooperation with IJCAI, Inc.

Co-sponsored by:
- AT&T Global Information Solutions
- NASA - Jet Propulsion Laboratory
- GTE

KDD-95 Conference Schedule

All KDD-95 sessions will be held in Room 408-BC, fourth level, Palais des Congrès de Montréal.

The KDD Opening Reception will be held in conjunction with Poster Session 1, Sunday, August 20, 6:00–8:00 PM, in Room 407A.

Conference Co chairs
Usama M. Fayyad, Jet Propulsion Lab,
California Institute of Technology
Ramasamy Uthurusamy, General Motors Research

Publicity Chair
Padhraic Smyth, Jet Propulsion Laboratory

Industry Liaison
Gregory Piatetsky-Shapiro,
GTE Laboratories

Demo Sessions Chair
Tej Anand, AT&T Global Information Solutions

Program Committee
Rakesh Agrawal, IBM Almaden Research Center, USA
Tej Anand, AT&T Global Information Solutions, USA
Ron Brachman, AT&T Bell Laboratories, USA
Wray Buntine, Heuristic Research, Inc., USA
Nick Cercone, University of Regina, Canada
Peter Cheeseman, NASA Ames Research Center, USA
Greg Cooper, University of Pittsburgh, USA

Brian Gaines, University of Calgary, Canada
Clark Glymour, Carnegie-Mellon University, USA
David Hand, Open University, UK
David Heckerman, Microsoft Corporation, USA
Se June Hong, IBM T. J. Watson Research Center, USA
Larry Jackel, AT&T Bell Laboratories, USA
Larry Kerschberg, George Mason University, USA
Willi Kloesgen, GMD, Germany
David Madigan, University of Washington, USA
Chris Matheus, GTE Laboratories, USA
Heikki Mannila, University of Helsinki, Finland
Gregory Piatetsky-Shapiro, GTE Laboratories, USA
Daryl Pregibon, AT&T Bell Laboratories, USA
Arno Siebes, CWI, The Netherlands
Evangelos Simoudis,
IBM Almaden Research Center, USA
Andrzej Skowron, University of Warsaw, Poland
Padhraic Smyth, Jet Propulsion Laboratory, USA
Alex Tuzhilin, NYU Stern School, USA
Xindong Wu, Monash University, Australia
Wojciech Ziarko, University of Regina, Canada
Jan Zytkow, Wichita State University, USA
Sunday, August 20, 1995

Room 408B, Palais des Congrès

8:30 - 9:00 AM

Welcome, Opening Remarks, Overview of KDD
Usama Fayyad, KDD–95 Cochair

9:00 - 10:15 AM

Session 1: Databases and Data Mining
Session Chair: Heikki Mannila

Applying a Data Miner To Heterogeneous Schema Integration
Son Duo and Brad Perry, Hughes Research Laboratories

Active Data Mining
Rakesh Agrawal and Giuseppe Psaila, IBM Almaden Research Center

A Database Interface for Clustering in Large Spatial Databases
Martin Ester, Hans-Peter Kriegel, and Xiaowei Xu, University of Munich, Germany

10:15 - 10:30 AM

Spotlight Session 1—Poster Summaries P1–P6

Paul Anderson, Ray Smith, and Zhongwei Zhang, Monash University, Australia

P2: Learning First Order Logic Rules with a Genetic Algorithm
S. Augier, G. Ventarini, and Y. Kodratoff, Univ: Paris-Sud, France

P3: Discovery and Maintenance of Functional Dependencies by Independencies
Siegfried Bell, University Dortmund, Germany

P4: Intelligent Instruments: Discovering How to Turn Spectral Data into Information
Wray L. Buntine and Tarang Patel, NASA Ames Research Center

P5: Designing Neural Networks from Statistical Models: A New Approach to Data Exploration
Antonio Caiani, McGill University, Canada and Yves Lechevallier INRIA-Rocquencourt, France

P6: Capacity and Complexity Control in Predicting the Spread Between Borrowing and Lending Interest Rates
Corinna Cortes, Harris Drucker, Dennis Hoover, and Vladimir Vapnik, AT&T Bell Laboratories

10:30 - 10:50 AM

Coffee Break

10:50 - 11:00 AM

Spotlight Session 2—Poster Summaries P7–P10

P7: Limits on Learning Machine Accuracy Imposed by Data Quality
Corinna Cortes, L. D. Jackel, and Wan-Ping Chiang, AT&T Bell Laboratories

P8: Knowledge Discovery in a Water Quality Database
Saso Dzeroski, Jozef Stefan Institute and Jasna Grbovic, Hydrometeorological Institute of Slovenia

P9: Data Mining for Loan Evaluation at ABN AMRO: A Case Study
A. J. Feelders and A. J. F. le Loux, University of Twente; J. W. van’t Zand, ABN AMRO Bank, The Netherlands

P10: Knowledge Discovery in Textual Databases (KDT)
Ronen Feldman and Ido Dagan, Bar-Ilan University, Israel

11:00 - 11:50 AM

Invited Speaker

Using Hidden Markov Models to Search Biosequence Databases
David Haussler, UCSC
11:50 AM - 12:00 PM  
**Spotlight Session 3: Poster Summaries P11–P14**

P11: Optimization and Simplification of Hierarchical Clusterings  
Doug Fisher, Vanderbilt University

P12: Structured and Unstructured Induction with EDAGs  
Brian R. Gaines, University of Calgary, Canada

P13: Restructuring Databases for Knowledge Discovery by Consolidation and Link Formation  
Henry G. Goldberg and Ted E. Senator, Financial Crimes Enforcement Network (FinCEN), U.S. Dept. of Treasury

P14: Rough Sets Similarity-Based Learning from Databases  
Xiaohua Hu and Nick Cercone, University of Regina, Canada

12:00 - 1:30 PM  
Lunch Break

1:30 - 2:30 PM  
**Panel Session**

Commercial KDD Applications: The Secret Ingredients for Success  
Panel Chairs: Gregory Pastek-Shapiro, GTE Laboratories and Evangelos Simoudis, IBM Almaden Research Center

2:30 - 3:20 PM  
**Session 2: Causality and Bayes Networks**

Session Chair: Alex Tuszilnik

Available Technology for Discovering Causal Models, Building Bayes Nets, and Selecting Predictors: The TETRAD II Program  
Clark Glymour, Carnegie Mellon University

Learning Bayesian Networks with Discrete Variables from Data  
Peter Spirtes and Christopher Meek, Carnegie Mellon University

3:20 - 3:30 PM  
**Spotlight Session 4: Poster Summaries P15–P17**

P15: Efficient Algorithms for Attribute-Oriented Induction  
Hoi-Yee Hwang and Wai-Chee Fu, Chinese University of Hong Kong

P16: Robust Decision Trees: Removing Outliers from Databases  
George H. John, Stanford University

P17: Conceptual Clustering in Structured Databases: A Practical Approach  
A. Ketron, P. Giannaros, and J. Kowczak, LSIT, Univ: Louis Pasteur, France

3:30 - 3:50 PM  
Coffee Break

3:50 - 5:00 PM  
**Parallel Session 3A: Rough Sets and Databases**

Room 408B, Palais des Congrès  
Session Chair: Jan Zytkow

Discovery of Concurrent Data Models from Experimental Tables: A Rough Set Approach  
Andrzej Skowron, Warsaw University and Zbigniew Suraj, Pedagogical University, Poland
Automated Discovery of Functional Components of Proteins from Amino-Acid Sequences Based on Rough Sets and Change of Representation
Shusaku Tsumoto and Hiroshi Tanaka, Tokyo Medical and Dental Univ., Japan

Using Rough Sets as Tools for Knowledge Discovery
Ning Shao, Wojciech Ziarko, Howard J. Hamilton, and Nick Cercone, University of Regina, Canada

Exploiting Upper Approximation in the Rough Set Methodology
Jitender S. Deogun, University of Nebraska at Lincoln; Vijay V. Raghavan and Hayri Sever, University of Southwestern Louisiana

A Perspective on Databases and Data Mining
Marcel Holstein and Martin Kersten, CWI Database Res. Group, The Netherlands; Heikki Mannila and Hannu Toivonen, University of Helsinki, Finland

Compression-Based Evaluation of Partial Determinations
Bernhard Pfahringer and Stefan Kramer, Austrian Research Inst. for AI, Austria

**Parallel Session 3B: Supervised Learning: Issues and Applications**

3:50 - 5:50 PM
Room 408C, Palais des Congrès
Session Chair: Willi Kloesgen

Knowledge Discovery in Telecommunication Services Data Using Bayesian Network Models
Kazuo J. Ezawa and Steve W. Norton, AT&T Bell Laboratories

Analyzing the Benefits of Domain Knowledge in Substructure Discovery
Sarvajani Dyke, Diane J. Cook, and Lawrence B. Holder, University of Texas at Arlington

Decision Tree Induction: How Effective is the Greedy Heuristic?
Sreerama K. Murthy and Steven Salzberg, Johns Hopkins University

Feature Subset Selection Using the Wrapper Method: Overfitting and Dynamic Search Space Topology
Ron Kohavi and Dan Sommerfield, Stanford University

Learning Arbiter and Combiner Trees from Partitioned Data for Scaling Machine Learning
Philip K. Chan and Salvatore J. Stifo, Columbia University

Are We Losing Accuracy While Gaining Confidence in Induced Rules: An Assessment of PrHL

**KDD-95 Opening Reception, Poster Session 1, and Demo Session**

6:00 - 8:00 PM
Room 407A, Palais des Congrès

**Poster Session One: Posters P1 through P17** (see handout)

**Demo Session**

Demo Session Chair: Tej Anand, AT&T Global Information Solutions

Knowledge Discovery from Multiple Databases
James Ribiero, George Mason University
Knowledge Discovery in Textual Databases
Ronen Feldman, Bar-Ilan University

Exploiting Visualization in Knowledge Discovery
Hung-Yan Lee, Huee-Leng Ong and Lee-Hian Quek, Information Technology Institute

KEFIR: The Key Findings Reporter for the Analysis of Healthcare Information
Christopher Mathews and Gregory Piatetsky-Shapiro, GTE Labs

Automated Large-scale Data Mining by Forty-Niner (49er)
Arun Sanjeev and Jan Zylkow

Monday, August 21, 1995

Room 408B, Palais des Congrès

8:30 - 8:40 AM

**Spotlight Session 5: Poster Summaries P18–P21**

P18: Anonymization Techniques for Knowledge Discovery in Databases
Willi Klosen, German National Research Center for Info. Technology (GMD)

P19: Exploiting Visualization in Knowledge Discovery
Hung-Yan Lee, Huee-Leng Ong, and Lee-Hian Quek, Information Technology Institute, Singapore

P20: Knowledge-Based Scientific Discovery in Geological Databases
Cen Li and Gaudam Biswas, Vanderbilt University

P21: An Iterative Improvement Approach for the Discretization of Numeric Attributes in Bayesian Classifiers
Michael J. Pazzani, University of California, Irvine

8:40 - 9:30 AM

**Session 4: Temporal Databases**

Session Chair: Wray Buntine

Fast Spatio-Temporal Data Mining of Large Geophysical Datasets
Paul Stolorz, JPL, et al.

Discovering Frequent Episodes in Sequences
H. Mannila, T. Toivonen, and A.J. Verkamo, Univ. of Helsinki

9:30 - 9:40 AM

**Spotlight Session 6: Poster Summaries P22–P25**

P22: Knowledge Discovery from Multiple Databases
James S. Ribeiro, Kenneth A. Kaufman, and Larry Kerschberg, George Mason University

P23: Discovering Enrollment Knowledge in University Databases
Arun P. Sanjeev and Jan M. Zylkow, Wichita State University

P24: Extracting Support Data for a Given Task
Bernhard Schoelkopf, Chris Burges, and Vladimir Vapnik, AT&T Bell Labs.

P25: Feature Extraction for Massive Data Mining
V. Sebadri and Raguram Sastikaran, AT&T Bell Laboratories; Shlom M. Weiss, Rutgers University

9:40 - 10:30 AM

**Invited Speaker**

A Database Perspective on Knowledge Discovery
Tomasz Imieliński, Rutgers University
10:30 - 10:50 AM  Coffee Break

10:50 - 11:00 AM  **Spotlight Session 7: Poster Summaries P26–P29**

**P26: Data Surveying: Foundations of an Inductive Query Language**  
Arno Siebes, CWI, Database Research Group, The Netherlands

**P27: On Subjective Measures of Interestingness in Knowledge Discovery**  
Avi Silberschatz, AT&T Bell Labs and Alexander Tzeitlin, New York Univ.

**P28: Using Recon for Data Cleaning**  
Evangelos Simoudis, IBM Almaden Research Center; Brian Livezey and Randy Kerber, Lockheed Palo Alto Research Laboratories

**P29: Accelerated Quantification of Bayesian Networks with Incomplete Data**  
Bo Thissin, Aalborg University, Denmark

11:00 - 11:50 AM  **Session 5: Inductive Learning**  
**Session Chair: Xindong Wu**

**MDL-Based Decision Tree Pruning**  
M. Mehta, J. Rissanen, and R. Agrawal, IBM Almaden Research Center

**Estimating the Robustness of Discovered Knowledge**  
Chun-Nam Hsu and Craig A. Knoblock, U.S.C.

11:50 AM - 12:00 PM  **Spotlight Session 8: Poster Summaries P30–P33**

**P30: Automated Selection of Rule Induction Methods Based on Recursive Iteration of Resampling Methods and Multiple Statistical Testing**  
Shusaku Tsumoto and Hiroshi Tanaka, Tokyo Medical and Dental Univ., Japan

**P31: Fuzzy Interpretation of Induction Results**  
Xindong Wu, Monash University; Australia and Petter Mablen, Royal Institute of Technology, Sweden

**P32: Resource and Knowledge Discovery in Global Information Systems: A Preliminary Design and Experiment**  
Osman R. Zaiane and Jewei Han, Simon Fraser University, Canada

**P33: Toward a Multi-Strategy and Cooperative Discovery System**  
Ning Zhong, The Univ. of Tokyo and Setsuo Ohsuga, The Waseda Univ., Japan

12:00 - 1:30 PM  Lunch Break

1:30 - 2:30 PM  **Invited Speaker**

Intelligent Local Learning: Statistical Algorithms for Prediction with High Dimensional Data  
Jerome Friedman, Stanford University

2:30 - 3:20 PM  **Session 6: KDD and Statistics**  
**Session Chair: Padhraic Smyth**

A Statistical Perspective on Knowledge Discovery In Databases  
John Elder, Rice Univ. and Daryl Pregibon, AT&T Bell Labs.

**Discriminant Adaptive Nearest Neighbor Classification**  
Trevor Hastie, Stanford University and Robert Tibshirani, University of Toronto

3:20 - 3:50 PM  Coffee Break

3:30 - 5:30 PM  **Poster Session 2: Posters P18 through P33** (see handout)  
Room 407A, Palais des Congrès
3:30 - 5:30 PM

**Demo Session (Repeated)**
Room 407A, Palais des Congrès

*Demo Session Chair: Tej Anand, AT&T Global Information Solutions*

Knowledge Discovery from Multiple Databases
*James Ribiero, George Mason University*

Knowledge Discovery in Textual Databases
*Ronen Feldman, Bar-Ilan University*

Exploiting Visualization in Knowledge Discovery
*Hing-Yan Lee, Hwee-Leng Ong and Lee-Hian Quek Information Technology Institute*

KEFIR: The Key Findings Reporter for the Analysis of Healthcare Information
*Christopher Mathews and Gregory Piatetsky-Shapiro, GTE Labs*

Automated Large-scale Data Mining by Forty-Niner (49er)
*Arun Sanjeev and Jan Zytkow*

5:30 - 6:00 PM

**Concluding Remarks, Summary and Wrap-up Session**
*R. Uthurusamy, Conference Cochair*
Room 408B, Palais des Congrès