# Table of Contents

## 2001 IEEE International Conference on Data Mining (ICDM 2001)

Welcome from the Steering Committee Chair .................................................. xiii
Message from the Program Chairs ................................................................. xv
Conference Organization ............................................................................... xvii
Steering Committee ...................................................................................... xviii
Program Committee ...................................................................................... xix
Non-PC Reviewers ......................................................................................... xxi

### Regular Papers

On Effective Conceptual Indexing and Similarity Search in Text Data .................. 3
  C. Aggarwal and P. Yu

Comparisons of Classification Methods for Screening Potential Compounds .......... 11
  A. An and Y. Wang

Knowledge Discovery from Diagrammatically Represented Data ....................... 19
  M. Anderson

Integrating E-Commerce and Data Mining: Architecture and Challenges ............. 27
  S. Ansari, R. Kohavi, L. Mason, and Z. Zheng

Classification with Degree of Membership: A Fuzzy Approach ......................... 35
  W.-H. Au and K. Chan

Provably Fast Training Algorithms for Support Vector Machines ...................... 43
  J. Balcazar, Y. Dai, and O. Watanabe

Who Links to Whom: Mining Linkage between Web Sites ................................ 51
  K. Bharat, B.-W. Chang, M. Henzinger, and M. Ruhl

Better Rules, Few Features: A Semantic Approach to Selecting Features from Text 59
  C. Blake and W. Pratt

Significance Tests for Patterns in Continuous Data .......................................... 67
  R. Bolton and D. Hand
Distributed Web Mining Using Bayesian Networks from Multiple Data Streams .................. 75  
R. Chen, K. Sivakumar, and H. Kargupta

A Hypergraph Based Clustering Algorithm for Spatial Data Sets .................................................. 83  
J.-S. Cherng and M.-J. Lo

Efficient Determination of Dynamic Split Points in a Decision Tree .............................................. 91  
D. Chickering, C. Meek, and R. Rounthwaite

Efficient Yet Accurate Clustering ................................................................................................... 99  
M. Dash, K. Tan, and H. Liu

A Min-max Cut Algorithm for Graph Partitioning and Data Clustering .............................................. 107  
C. Ding, X. He, H. Zha, M. Gu, and H. Simon

Preprocessing Opportunities in Optimal Numerical Range Partitioning ...................................... 115  
T. Elomaa and J. Rousu

Using Artificial Anomalies to Detect Unknown and Known Network Intrusions ...................... 123  
W. Fan, M. Miller, S. Stolfo, W. Lee, and P. Chan

Using Rule Sets to Maximize ROC Performance ........................................................................... 131  
T. Fawcett

A Synchronization Based Algorithm for Discovering Ellipsoidal Clusters in Large Datasets ........ 139  
H. Frigui and M. Rhouma

Functional Trees for Classification ............................................................................................... 147  
J. Gama

A Tight Upper Bound on the Number of Candidate Patterns ...................................................... 155  
F. Geerts, B. Goethals, and J. Van den Bussche

Efficiently Mining Maximal Frequent Itemsets ............................................................................. 163  
K. Gouda and M. Zaki

The DIAsDEM Framework for Converting Domain-Specific Texts into XML Documents with Data Mining Techniques ................................................................. 171  
H. Graubitz, M. Spiliopoulou, and K. Winkler

A Scalable Algorithm for Clustering Sequential Data ................................................................ 179  
V. Guralnik and G. Karypis

Clustering Validity Assessment: Finding the Optimal Partitioning of a Data Set .......................... 187  
M. Halkidi and M. Vazirgiannis
Automatic Topic Identification Using Webpage Clustering .......................................................... 195
   X. He, C. Ding, H. Zha, and H. Simon

Time Series Segmentation for Context Recognition in Mobile Devices .................................... 203
   J. Himberg, K. Korpiaho, H. Mannila, J. Tikanmäki, and H. Toivonen

Indiscernibility Degree of Objects for Evaluating Simplicity of Knowledge
in the Clustering Procedure .......................................................................................................... 211
   S. Hirano and S. Tsumoto

Mining Coverage-Based Fuzzy Rules by Evolutional Computation ........................................... 218
   T.-P. Hong and Y.-C. Lee

An Efficient Fuzzy C-Means Clustering Algorithm .................................................................... 225
   M.-C. Hung and D.-L. Yang

Using Rough Sets Theory and Database Operations to Construct a Good
Ensemble of Classifiers for Data Mining Applications ................................................................. 233
   X. Hu

Fuzzy Data Mining: Effect of Fuzzy Discretization ...................................................................... 241
   H. Ishibuchi, T. Yamamoto, and T. Nakashima

The Computational Complexity of High-Dimensional Correlation Search .................................. 249
   C. Jermaine

Evaluating Boosting Algorithms to Classify Rare Classes:
Comparison and Improvements ................................................................................................... 257
   M. Joshi, V. Kumar, and R. Agarwal

An Agglomerative Hierarchical Clustering Using Partial Maximum Array
and Incremental Similarity Computation Method ......................................................................... 265
   S. Jung and T.-S. Kim

Distance Measures for Effective Clustering of ARIMA Time-Series ............................................ 273
   K. Kalpakis, D. Gada, and V. Puttagunta

Mining Decision Trees from Data Streams in a Mobile Environment ......................................... 281
   H. Kargupta and B.-H. Park

An Online Algorithm for Segmenting Time Series .................................................................... 289
   E. Keogh, S. Chu, D. Hart, and M. Pazzani

AINE: An Immunological Approach to Data Mining ..................................................................... 297
   T. Knight and J. Timmis
Concise Representation of Frequent Patterns Based on Disjunction-Free Generators.............. 305
M. Kryszkiewicz

Frequent Subgraph Discovery ..................................................................................................... 313
M. Kuramochi and G. Karypis

Statistical Considerations in Learning from Data ........................................................................ 321
H. Kyburg, Jr.

Subject Classification in the Oxford English Dictionary ............................................................... 329
Z. Langari and F. Tompa

On Mining General Temporal Association Rules in a Publication Database .............................. 337
C.-H. Lee, C.-R. Lin, and M.-S. Chen

Preparations for Semantics-Based XML Mining ......................................................................... 345
J.-W. Lee, K. Lee, and W. Kim

Mining Image Features for Efficient Query Processing ............................................................... 353
B. Li, W.-C. Lai, E. Chang, and K.-T. Cheng

Mining the Smallest Association Rule Set for Predictions ........................................................ 361
J. Li, H. Shen, and R. Topor

CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules .......... 369
W. Li, J. Han, and J. Pei

Analyzing the Interestingness of Association Rules from the Temporal Dimension ............... 377
B. Liu, Y. Ma, and R. Lee

Closing the Loop: An Agenda- and Justification-Based Framework for Selecting the Next Discovery Task to Perform ......................................................... 385
G. Livingston, J. Rosenberg, and B. Buchanan

Closing the Loop: Heuristics for Autonomous Discovery .......................................................... 393
G. Livingston, J. Rosenberg, and B. Buchanan

Anchor Text Mining for Translation of Web Queries ............................................................. 401
W.-H. Lu, L.-F. Chien, and H.-J. Lee

Mining Mutually Dependent Patterns ...................................................................................... 409
S. Ma and J. Hellerstein

The EQ Framework for Learning Equivalence Classes of Bayesian Networks ..................... 417
P. Munteanu and M. Bendou
Visualizing Association Mining Results through Hierarchical Clusters ........................................ 425
  S. Noel, V. Raghavan, and C.-H. Chu

Mining Constrained Association Rules to Predict Heart Disease ................................................ 433
  C. Ordonez, E. Omiecinski, L. de Braal, C. Santana, N. Ezquerra,
  J. Taboada, D. Cooke, E. Krawczynska, and E. Garcia

H-Mine: Hyper-Structure Mining of Frequent Patterns in Large Databases .................................. 441
  J. Pei, J. Han, H. Lu, S. Nishio, S. Tang, and D. Yang

FARM: A Framework for Exploring Mining Spaces with Multiple Attributes ................................ 449
  C.-S. Perng, H. Wang, S. Ma, and J. Hellerstein

Neural Analysis of Mobile Radio Access Network ........................................................................ 457
  K. Raivio, O. Simula, and J. Laiho

Discovery of Association Rules in Tabular Data ........................................................................... 465
  G. Richards and V. Rayward-Smith

Theory and Applications of Attribute Decomposition .................................................................. 473
  L. Rokach and O. Mainon

FiExPat: Flexible Extraction of Sequential Patterns ...................................................................... 481
  P.-Y. Rolland

Interestingness PreProcessing ........................................................................................................ 489
  S. Sahar

Data Analysis and Mining in Ordered Information Tables .......................................................... 497
  Y. Sai, Y. Yao, and N. Zhong

LPMiner: An Algorithm for Finding Frequent Itemsets Using Length-Decreasing Support Constraint 505
  M. Seno and G. Karypis

Document Clustering and Cluster Topic Extraction in Multilingual Corpora ................................ 513
  J. Silva, J. Mexia, A. Coelho, and G. Lopes

Hierarchical Text Classification and Evaluation .......................................................................... 521
  A. Sun and E.-P. Lim

Web Cartography for Online State Promotion: An Algorithm for Clustering Web Resources .......... 529
  F. Velin, P. Kuntz, and H. Briand
Maintenance of Sequential Patterns for Record Deletion ............................................................ 536
C.-Y. Wang, T.-P. Hong, and S.-S. Tseng

SSDT: A Scalable Subspace-Splitting Classifier for Biased Data ............................................... 542
H. Wang and P. Yu

Meta-patterns: Revealing Hidden Periodic Patterns ..................................................................... 550
W. Wang, J. Yang, and P. Yu

Using Boosting to Simplify Classification Models ........................................................................ 558
V. Wheway

Interestingness, Peculiarity, and Multi-database Mining ............................................................. 566
N. Zhong, Y. Yao, M. Ohshima, and S. Ohsuga

Posters

Discovering Similar Patterns for Characterising Time Series in a Medical Domain .................... 577
F. Alonso, J. Caraça-Valente, L. Martínez, and C. Montes

Creating Ensembles of Classifiers ............................................................................................... 580
N. Chawla, S. Eschrich, and L. Hall

Association Rules Enhanced Classification of Underwater Acoustic Signal ............................... 582
J. Chen, H. Li, and S. Tang

Efficient Splitting Rules Based on the Probabilities of Pre-assigned Intervals ............................ 584
J.-S. Cho and N. Adam

Inexact Field Learning: An Approach to Induce High Quality Rules from Low Quality Data ........ 586
H. Dai, X. Hang, and G. Li

Incremental Support Vector Machine Construction ..................................................................... 589
C. Domeniconi and D. Gunopulos

Mining Generalized Association Rules for Sequential and Path Data ......................................... 593
W. Gaul and L. Schmidt-Thieme

Combining Labeled and Unlabeled Data for Text Classification with a Large Number of Categories .......................... 597
R. Ghani

Dependency Derivation in Industrial Process Data ...................................................................... 599
D. Gillblad and A. Holst
Discovering Representative Episodal Association Rules from Event Sequences Using Frequent Closed Episode Sets and Event Constraints ............................................. 603
    S. Harms, J. Deogun, J. Saquer, and T. Tadesse

Text Clustering Based on Good Aggregations .............................................. 607
    A. Hotho, A. Maedche, and S. Staab

Ad Hoc Association Rule Mining as SQL3 Queries ........................................... 609
    H. Jamil

Heuristic Optimization for Decentralized Frequent Itemset Counting .................. 613
    V. Jensen and N. Soparkar

Evolutionary Structure Learning Algorithm for Bayesian Network and Penalized Mutual Information Metric ................................................................. 615
    G. Li, F. Tong, and H. Dai

Applications of Data Mining in Hydrology ....................................................... 617
    X. Liang and Y. Liang

RPCL-Based Local PCA Algorithm ................................................................. 621
    Z. Liu and L. Xu

Learning Automatic Acquisition of Subcategorization Frames Using Bayesian Inference and Support Vector Machines ......................................................... 623
    M. Maragoudakis, K. Kermanidis, N. Fakotakis, and G. Kokkinakis

Bayesian Data Mining on the Web with B-Course ............................................. 626
    P. Myllymäki, T. Silander, H. Tirri, and P. Uronen

An Experimental Comparison of Supervised and Unsupervised Approaches to Text Summarization ................................................................. 630
    T. Nomoto and Y. Matsumoto

A Fast Algorithm to Cluster High Dimensional Basket Data ............................ 633
    C. Ordonez, E. Omiecinski, and N. Ezquerra

Metric Rule Generation with Septic Shock Patient Data .................................... 637
    J. Paetz

The Representative Basis for Association Rules ................................................ 639
    V. Phan-Luong

Incremental Learning with Support Vector Machines ....................................... 641
    S. Rüping
A Clustering Method for Very Large Mixed Data Sets ......................................................... 643
   G. Sánchez-Díaz and J. Ruiz-Shulcloper

Mining the Web with Active Hidden Markov Models .......................................................... 645
   T. Scheffer, C. Decomain, and S. Wrobel

A Simple KNN Algorithm for Text Categorization .............................................................. 647
   P. Soucy and G. Mineau

Measuring Real-Time Predictive Models .............................................................................. 649
   S. Steingold, R. Wherry, and G. Piatetsky-Shapiro

Incremental Learning of Bayesian Networks with Hidden Variables ................................... 651
   F. Tian, H. Zhong, Y. Lu, and C. Shi

Mining Frequent Closed Itemsets with the Frequent Pattern List ........................................... 653
   F.-C. Tseng, C.-C. Hsu, and H. Chen

Classification through Maximizing Density ........................................................................ 655
   H. Wang, I. Düntsch, D. Bell, and D. Liu

An Immune Neural Network Used for Classification .............................................................. 657
   L. Wang and L. Jiao

α-Surface and Its Application to Mining Protein Data ............................................................. 659
   X. Wang

An Efficient Data Mining Technique for Discovering Interesting Sequential Patterns .......... 663
   S.-J. Yen and Y.-S. Lee

Fast Parallel Association Rule Mining without Candidacy Generation ................................... 665
   O. Zaiane, M. El-Hajj, and P. Lu

A Comparison of Stacking with Meta Decision Trees to Bagging, Boosting, and Stacking with other Methods ................................................................. 669
   B. Ženko, L. Todorovski, and S. Džeroski

Mining California Vital Statistics Data ..................................................................................... 671
   A. Zhang, Q. Ha, and M. Lu

A Pattern Decomposition (PD) Algorithm for Finding All Frequent Patterns in Large Datasets ................................................................. 673
   Q. Zou, W. Chu, D. Johnson, and H. Chiu

Author Index .......................................................................................................................... 675