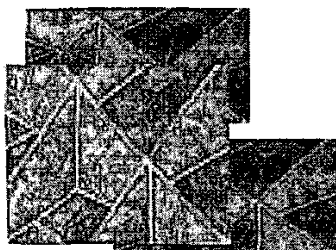


Proceedings

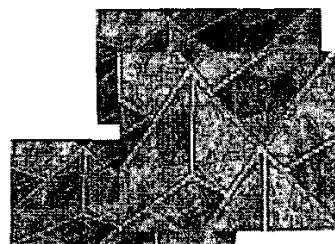


2002 IEEE International Conference on Data Mining

ICDM 2002



Proceedings



2002 IEEE International Conference on Data Mining

ICDM 2002

9-12 December 2002 • Maebashi City, Japan

Editors

Vipin Kumar, University of Minnesota, USA
Shusaku Tsumoto, Shimane Medical University, Japan
Ning Zhong, Maebashi Institute of Technology, Japan
Philip S. Yu, IBM T. J. Watson Research Center, USA
Xindong Wu, University of Vermont, USA

Sponsored by

IEEE Computer Society Technical Committee on
Pattern Analysis and Machine Intelligence (TCPAMI)
IEEE Computer Society Technical Committee on Computational Intelligence (TCCI)



IEEE
COMPUTER
SOCIETY

Los Alamitos, California

Washington • Brussels • Tokyo

Copyright © 2002 by The Institute of Electrical and
Electronics Engineers, Inc.
All rights reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 133, Piscataway, NJ 08855-1331.

The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society, or the Institute of Electrical and Electronics Engineers, Inc.

IEEE Computer Society Order Number PR01754

ISBN 0-7695-1754-4

Library of Congress Number 2002109464

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1-714-821-8380
Fax: + 1-714-821-4641
E-mail: cs.books@computer.org

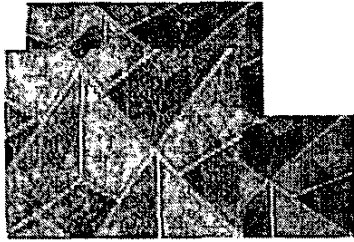
IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1-732-981-0060
Fax: + 1-732-981-9667
[http://shop.ieee.org/store/
customer-service@ieee.org](http://shop.ieee.org/store/customer-service@ieee.org)

IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: + 81-3-3408-3118
Fax: + 81-3-3408-3553
tokyo.ofc@computer.org

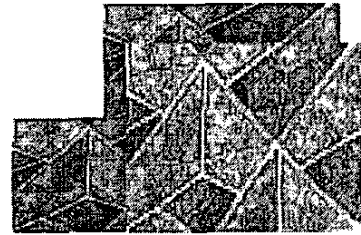
Editorial production by Stephanie Kawada
Cover art production by Joe Daigle/Studio Productions
Printed in the United States of America by The Printing House


IEEE
COMPUTER
SOCIETY





Proceedings



ICDM 2002

Table of Contents

Message from the Conference Chairs and Program Chairs	xiv
Conference Organization	xvi
Steering Committee	xvii
Program Committee	xviii
Non-PC Reviewers	xxi
Corporate Sponsors	xxii
Tutorials	783
Workshops	784

Main-Track Regular Papers

Empirical Comparison of Various Reinforcement Learning Strategies for Sequential Targeted Marketing <i>N. Abe, E. Pednault, H. Wang, B. Zadrozny, W. Fan, and C. Apte</i>	3
Investigative Profiling with Computer Forensic Log Data and Association Rules <i>T. Abraham and O. de Vel</i>	11

Text Document Categorization by Term Association.....	19
<i>M.-L. Antonie and O. R. Zaïane</i>	
Online Algorithms for Mining Semi-Structured Data Stream.....	27
<i>T. Asai, H. Arimura, K. Abe, S. Kawasoe, and S. Arikawa</i>	
A Lazy Approach to Pruning Classification Rules.....	35
<i>E. Baralis and P. Garza</i>	
High Performance Data Mining Using the Nearest Neighbor Join.....	43
<i>C. Böhm and F. Krebs</i>	
Mining Molecular Fragments: Finding Relevant Substructures of Molecules.....	51
<i>C. Borgelt and M. R. Berthold</i>	
Mining General Temporal Association Rules for Items with Different Exhibition Periods.....	59
<i>C.-Y. Chang, M.-S. Chen, and C.-H. Lee</i>	
Learning with Progressive Transductive Support Vector Machine.....	67
<i>Y. Chen, G. Wang, and S. Dong</i>	
Towards Automatic Generation of Query Taxonomy: A Hierarchical Query Clustering Approach.....	75
<i>S.-L. Chuang and L.-F. Chien</i>	
Evolutionary Time Series Segmentation for Stock Data Mining.....	83
<i>F. Chung, T. Fu, R. Luk, and V. Ng</i>	
Using Functional PCA for Cardiac Motion Exploration.....	91
<i>D. Clot</i>	
Unsupervised Segmentation of Categorical Time Series into Episodes.....	99
<i>P. Cohen, B. Heeringa, and N. Adams</i>	
Speed-Up Iterative Frequent Itemset Mining with Constraint Changes.....	107
<i>G. Cong and B. Liu</i>	
Feature Selection for Clustering—A Filter Solution.....	115
<i>M. Dash, K. Choi, P. Scheuermann, and H. Liu</i>	
A Theory of Inductive Query Answering.....	123
<i>L. De Raedt, M. Jaeger, S. D. Lee, and H. Mannila</i>	
Iterative Clustering of High Dimensional Text Data Augmented by Local Search.....	131
<i>I. S. Dhillon, Y. Guan, and J. Kogan</i>	

Cluster Merging and Splitting in Hierarchical Clustering Algorithms	139
<i>C. Ding and X. He</i>	
Adaptive Dimension Reduction for Clustering High Dimensional Data	147
<i>C. Ding, X. He, H. Zha, and H. D. Simon</i>	
Modal-Style Operators in Qualitative Data Analysis.....	155
<i>I. Düntsch and G. Gediga</i>	
Progressive Modeling	163
<i>W. Fan, H. Wang, P. S. Yu, S. Lo, and S. Stolfo</i>	
Unsupervised Clustering of Symbol Strings and Context Recognition	171
<i>J. A. Flanagan, J. Mäntylarvi, and J. Himberg</i>	
A Parameterless Method for Efficiently Discovering Clusters of Arbitrary Shape in Large Datasets.....	179
<i>A. Foss and O. R. Zaïane</i>	
Discriminative Category Matching: Efficient Text Classification for Huge Document Collections	187
<i>G. P. C. Fung, J. X. Yu, and H. Lu</i>	
Using Text Mining to Infer Semantic Attributes for Retail Data Mining	195
<i>R. Ghani and A. E. Fano</i>	
Phrase-Based Document Similarity Based on an Index Graph Model	203
<i>K. M. Hammouda and M. S. Kamel</i>	
Mining Top-K Frequent Closed Patterns without Minimum Support	211
<i>J. Han, J. Wang, Y. Lu, and P. Tzvetkov</i>	
Mining Similar Temporal Patterns in Long Time-Series Data and Its Application to Medicine	219
<i>S. Hirano and S. Tsumoto</i>	
Mining Generalized Association Rules Using Pruning Techniques	227
<i>Y.-F. Huang and C.-M. Wu</i>	
A Formal Model for User Preference	235
<i>S. Y. Jung, J.-H. Hong, and T.-S. Kim</i>	
Convex Hull Ensemble Machine	243
<i>Y. Kim</i>	
Recognition of Common Areas in a Web Page Using Visual Information: A Possible Application in a Page Classification	250
<i>M. Kovacevic, M. Diligenti, M. Gori, and V. Milutinovic</i>	
Discovering Frequent Geometric Subgraphs	258
<i>M. Kuramochi and G. Karypis</i>	

Adapting Classification Rule Induction to Subgroup Discovery	266
<i>N. Lavrač, P. Flach, B. Kavšek, and L. Todorovski</i>	
Linear Causal Model Discovery Using the MML Criterion	274
<i>G. Li, H. Dai, and Y. Tu</i>	
Attribute (Feature) Completion—The Theory of Attributes from Data Mining Prospect.....	282
<i>T. Y. Lin</i>	
O-Cluster: Scalable Clustering of Large High Dimensional Data Sets.....	290
<i>B. L. Milenova and M. M. Campos</i>	
Employing Discrete Bayes Error Rate for Discretization and Feature Selection Tasks	298
<i>A. Mittal and L.-F. Cheong</i>	
Feature Selection Algorithms: A Survey and Experimental Evaluation.....	306
<i>L. C. Molina, L. Belanche, and À. Nebot</i>	
Multivariate Supervised Discretization, A Neighborhood Graph Approach	314
<i>F. Muhlenbach and R. Rakotomalala</i>	
Mining Association Rules from Stars	322
<i>E. K. K. Ng, A. W.-C. Fu, and K. Wang</i>	
PERUSE: An Unsupervised Algorithm for Finding Recurring Patterns in Time Series.....	330
<i>T. Oates</i>	
Adaptive and Resource-Aware Mining of Frequent Sets.....	338
<i>S. Orlando, P. Palmerini, R. Perego, and F. Silvestri</i>	
A New Implementation Technique for Fast Spectral Based Document Retrieval Systems	346
<i>L. A. F. Park, M. Palaniswami, and K. Ramamohanarao</i>	
Efficient Progressive Sampling for Association Rules	354
<i>S. Parthasarathy</i>	
Efficient Discovery of Common Substructures in Macromolecules.....	362
<i>S. Parthasarathy and M. Coatney</i>	
Mining Motifs in Massive Time Series Databases	370
<i>P. Patel, E. Keogh, J. Lin, and S. Lonardi</i>	
On Computing Condensed Frequent Pattern Bases	378
<i>J. Pei, G. Dong, W. Zou, and J. Han</i>	
Automatic Web Page Classification in a Dynamic and Hierarchical Way	386
<i>X. Peng and B. Choi</i>	

User-Directed Exploration of Mining Space with Multiple Attributes	394
<i>C.-S. Perng, H. Wang, S. Ma, and J. L. Hellerstein</i>	
Mining Significant Associations in Large Scale Text Corpora.....	402
<i>P. Raghavan and P. Tsaparas</i>	
On a Capacity Control Using Boolean Kernels for the Learning of Boolean Functions	410
<i>K. Sadohara</i>	
SLPMiner: An Algorithm for Finding Frequent Sequential Patterns Using Length-Decreasing Support Constraint	418
<i>M. Seno and G. Karypis</i>	
Objective-Oriented Utility-Based Association Mining.....	426
<i>Y.-D. Shen, Z. Zhang, and Q. Yang</i>	
A Self-Organizing Map with Expanding Force for Data Clustering and Visualization	434
<i>W.-H. Shum, H.-D. Jin, K.-S. Leung, and M.-L. Wong</i>	
On the Mining of Substitution Rules for Statistically Dependent Items.....	442
<i>W.-G. Teng, M.-J. Hsieh, and M.-S. Chen</i>	
TreeFinder: A First Step towards XML Data Mining.....	450
<i>A. Termier, M.-C. Rousset, and M. Sebag</i>	
Computing Frequent Graph Patterns from Semistructured Data	458
<i>N. Vanetik, E. Gudes, and S. E. Shimony</i>	
Estimating the Number of Segments in Time Series Data Using Permutation Tests	466
<i>K. T. Vasko and H. T. T. Toivonen</i>	
Predicting Rare Events in Temporal Domains.....	474
<i>R. Vilalta and S. Ma</i>	
Mining Associations by Pattern Structure in Large Relational Tables	482
<i>H. Wang, C.-S. Perng, S. Ma, and P. S. Yu</i>	
Comparison of Lazy Bayesian Rule and Tree-Augmented Bayesian Learning.....	490
<i>Z. Wang and G. I. Webb</i>	
A Hybrid Approach to Discover Bayesian Networks from Databases Using Evolutionary Programming	498
<i>M. L. Wong, S. Y. Lee, and K. S. Leung</i>	
Adapting Information Extraction Knowledge For Unseen Web Sites.....	506
<i>T.-L. Wong and W. Lam</i>	

From Path Tree to Frequent Patterns: A Framework for Mining Frequent Patterns	514
<i>Y. Xu, J. X. Yu, G. Liu, and H. Lu</i>	
Mining Case Bases for Action Recommendation	522
<i>Q. Yang and H. Cheng</i>	
Adaptive Ripple Down Rules Method Based on Minimum Description Length Principle.....	530
<i>T. Yoshida, T. Wada, H. Motoda, and T. Washio</i>	
Heterogeneous Learner for Web Page Classification	538
<i>H. Yu, K. C.-C. Chang, and J. Han</i>	
Using Category-Based Adherence to Cluster Market-Basket Data	546
<i>C.-H. Yun, K.-T. Chuang, and M.-S. Chen</i>	
A Comparison Study on Algorithms for Incremental Update of Frequent Sequences	554
<i>M. Zhang, B. Kao, and C.-L. Yip</i>	
On Active Learning for Data Acquisition.....	562
<i>Z. Zheng and B. Padmanabhan</i>	
SmartMiner: A Depth First Algorithm Guided by Tail Information for Mining Maximal Frequent Itemsets	570
<i>Q. Zou, W. W. Chu, and B. Lu</i>	

Main-Track Short Papers

Neighborgram Clustering: Interactive Exploration of Cluster Neighborhoods	581
<i>M. R. Berthold, B. Wiswedel, and D. E. Patterson</i>	
A New Algorithm for Learning Parameters of a Bayesian Network from Distributed Data	585
<i>R. Chen and K. Sivakumar</i>	
Optimal Projections of High Dimensional Data	589
<i>E. Corchado and C. Fyfe</i>	
Generating an Informative Cover for Association Rules.....	597
<i>L. Cristofor and D. Simovici</i>	
Extraction Techniques for Mining Services from Web Sources.....	601
<i>H. Davulcu, S. Mukherjee, and I. V. Ramakrishnan</i>	
Reviewing RELIEF and Its Extensions: A New Approach for Estimating Attributes Considering High-Correlated Features	605
<i>R. Flórez-López</i>	

ESRS: A Case Selection Algorithm Using Extended Similarity-Based Rough Sets	609
<i>L. Geng and H. J. Hamilton</i>	
An Algebraic Approach to Data Mining: Some Examples	613
<i>R. L. Grossman and R. G. Larson</i>	
Wavelet Based UXO Detection	617
<i>S. Hodgson, N. Dunstan, and R. Murison</i>	
Ensemble Modeling through Multiplicative Adjustment of Class Probability	621
<i>S. J. Hong, J. Hosking, and R. Natarajan</i>	
Mining a Set of Coregulated RNA Sequences	625
<i>Y.-J. Hu</i>	
Association Analysis with One Scan of Databases	629
<i>H. Huang, X. Wu, and R. Relue</i>	
Implementation of a Least Fixpoint Operator for Fast Mining of Relational Databases	633
<i>H. M. Jamil</i>	
Considering Both Intra-Pattern and Inter-Pattern Anomalies for Intrusion Detection	637
<i>N. Jiang, K. A. Hua, and S. Sheu</i>	
On Evaluating Performance of Classifiers for Rare Classes	641
<i>M. V. Joshi</i>	
Learning from Order Examples	645
<i>T. Kamishima and S. Akaho</i>	
A Personalized Music Filtering System Based on Melody Style Classification	649
<i>F.-F. Kuo and M.-K. Shan</i>	
Solving the Fragmentation Problem of Decision Trees by Discovering Boundary Emerging Patterns	653
<i>J. Li and L. Wong</i>	
Improving Medical/Biological Data Classification Performance by Wavelet Preprocessing	657
<i>Q. Li, T. Li, S. Zhu, and C. Kambhamettu</i>	
Progressive and Interactive Analysis of Event Data Using Event Miner	661
<i>S. Ma, J. L. Hellerstein, C. Perng, and G. Grabarnik</i>	
Toward XML-Based Knowledge Discovery Systems	665
<i>R. Meo and G. Psaila</i>	

Using Sequential and Non-Sequential Patterns in Predictive Web Usage Mining Tasks	669
<i>B. Mobasher, H. Dai, T. Luo, and M. Nakagawa</i>	
Intersection Based Generalization Rules for the Analysis of Symbolic Septic Shock Patient Data	673
<i>J. Paetz</i>	
Exploring Interestingness through Clustering: A Framework	677
<i>S. Sahar</i>	
On Incorporating Subjective Interestingness into the Mining Process	681
<i>S. Sahar</i>	
Exploring the Parameter State Space of Stacking	685
<i>A. K. Seewald</i>	
Mining Associated Implication Networks: Computational Intermarket Analysis	689
<i>P. Tse and J. Liu</i>	
Maintenance of Sequential Patterns for Record Modification Using Pre-Large Sequences	693
<i>C.-Y. Wang, T.-P. Hong, and S.-S. Tseng</i>	
Concept Tree Based Clustering Visualization with Shaded Similarity Matrices	697
<i>J. Wang, B. Yu, and L. Gasser</i>	
ΔB^+ Tree: Indexing 3D Point Sets for Pattern Discovery	701
<i>X. Wang</i>	
An Incremental Approach to Building a Cluster Hierarchy	705
<i>D. H. Widyantoro, T. R. Ioerger, and J. Yen</i>	
A Comparative Study of RNN for Outlier Detection in Data Mining	709
<i>G. Williams, R. Baxter, H. He, S. Hawkins, and L. Gu</i>	
Evaluating the Utility of Statistical Phrases and Latent Semantic Indexing for Text Classification	713
<i>H. Wu and D. Gunopulos</i>	
Mixtures of ARMA Models for Model-Based Time Series Clustering	717
<i>Y. Xiong and D.-Y. Yeung</i>	
gSpan: Graph-Based Substructure Pattern Mining	721
<i>X. Yan and J. Han</i>	
InfoMiner+: Mining Partial Periodic Patterns with Gap Penalties	725
<i>J. Yang, W. Wang, and P. S. Yu</i>	

FD_Mine: Discovering Functional Dependencies in a Database Using Equivalences	729
<i>H. Yao, H. J. Hamilton, and C. J. Butz</i>	
Mining Genes in DNA Using GeneScout	733
<i>M. M. Yin and J. T. L. Wang</i>	
Clustering Spatial Data when Facing Physical Constraints	737
<i>O. R. Zaïane and C.-H. Lee</i>	
Mining Surveillance Video for Independent Motion Detection.....	741
<i>Z. Zhang</i>	
Adaptive Parallel Sentences Mining from Web Bilingual News Collection	745
<i>B. Zhao and S. Vogel</i>	

Industry-Track Papers

Telecommunications Strategic Marketing—KDD and Economic Modeling	751
<i>S. Cazzella, L. Dragone, and S. M. Trisolini</i>	
webSPADE: A Parallel Sequence Mining Algorithm to Analyze Web Log Data	755
<i>A. Demiriz</i>	
Mining Online Users' Access Records for Web Business Intelligence	759
<i>S. Fong and S. Chan</i>	
Discovery of Interesting Association Rules from Livelink Web Log Data	763
<i>X. Huang, A. An, N. Cercone, and G. Promhouse</i>	
Mining Optimal Actions for Profitable CRM	767
<i>C. X. Ling, T. Chen, Q. Yang, and J. Cheng</i>	
Visually Mining Web User Clickpaths	771
<i>T. Mah and Y. Li</i>	
Experimentation and Self Learning in Continuous Database Marketing.....	775
<i>J. E. Pearce, R. N. Shaw, G. I. Webb, and B. Garner</i>	
Demand Forecasting by the Neural Network with Discrete Fourier Transform	779
<i>M. Yohda, M. Saito-Arita, A. Okada, R. Suzuki, and Y. Kakemoto</i>	

Author Index	785
---------------------------	-----