

# Table of Contents

## Session 1A – Time Series

Scaling up Dynamic Time Warping to Massive Dataset . . . . .	1
<i>E.J. Keogh, M.J. Pazzani</i>	
The Haar Wavelet Transform in the Time Series Similarity Paradigm . . . . .	12
<i>Z.R. Struzik, A. Siebes</i>	
Rule Discovery in Large Time-Series Medical Databases . . . . .	23
<i>S. Tsumoto</i>	

## Session 1B – Applications

Simultaneous Prediction of Multiple Chemical Parameters of River Water Quality with TILDE . . . . .	32
<i>H. Blockeel, S. Džeroski, J. Grbović</i>	
Applying Data Mining Techniques to Wafer Manufacturing . . . . .	41
<i>E. Bertino, B. Catania, E. Caglio</i>	
An Application of Data Mining to the Problem of the University Students’ Dropout Using Markov Chains . . . . .	51
<i>S. Massa, P.P. Puliafito</i>	

## Session 2A – Taxonomies and Partitions

Discovering and Visualizing Attribute Associations Using Bayesian Networks and Their Use in KDD . . . . .	61
<i>G. Masuda, R. Yano, N. Sakamoto, K. Ushijima</i>	
Taxonomy Formation by Approximate Equivalence Relations, Revisited . . .	71
<i>F.A. El-Mouadib, J. Koronacki, J.M. Żytkow</i>	
On the Use of Self-Organizing Maps for Clustering and Visualization . . . . .	80
<i>A. Flexer</i>	
Speeding Up the Search for Optimal Partitions . . . . .	89
<i>T. Elomaa, J. Rousu</i>	

## Session 2B – Logic Methods

Experiments in Meta-level Learning with ILP . . . . .	98
<i>L. Todorovski, S. Džeroski</i>	

Boolean Reasoning Scheme with Some Applications in Data Mining . . . . .	107
<i>A. Skowron, H.S. Nguyen</i>	
On the Correspondence between Classes of Implicational and Equivalence Quantifiers . . . . .	116
<i>J. Ivánek</i>	
Querying Inductive Databases via Logic-Based User-Defined Aggregates . .	125
<i>F. Giannotti, G. Manco</i>	

## Session 3A – Distributed and Multirelational Databases

Peculiarity Oriented Multi-database Mining . . . . .	136
<i>N. Zhong, Y.Y. Yao, S. Ohsuga</i>	
Knowledge Discovery in Medical Multi-databases: A Rough Set Approach .	147
<i>S. Tsumoto</i>	
Automated Discovery of Rules and Exceptions from Distributed Databases Using Aggregates . . . . .	156
<i>R. Páircéir, S. McClean, B. Scotney</i>	

## Session 3B – Text Mining and Feature Selection

Text Mining via Information Extraction . . . . .	165
<i>R. Feldman, Y. Aumann, M. Fresko, O. Liphstat, B. Rosenfeld, Y. Schler</i>	
TopCat: Data Mining for Topic Identification in a Text Corpus . . . . .	174
<i>C. Clifton, R. Cooley</i>	
Selection and Statistical Validation of Features and Prototypes . . . . .	184
<i>M. Sebban, D.A. Zighed, S. Di Palma</i>	

## Session 4A – Rules and Induction

Taming Large Rule Models in Rough Set Approaches . . . . .	193
<i>T. Ágotnes, J. Komorowski, T. Løken</i>	
Optimizing Disjunctive Association Rules . . . . .	204
<i>D. Zelenko</i>	
Contribution of Boosting in Wrapper Models . . . . .	214
<i>M. Sebban, R. Nock</i>	
Experiments on a Representation-Independent "Top-Down and Prune" Induction Scheme . . . . .	223
<i>R. Nock, M. Sebban, P. Jappy</i>	

## Session 5A – Interesting and Unusual

Heuristic Measures of Interestingness .....	232
<i>R.J. Hilderman, H.J. Hamilton</i>	
Enhancing Rule Interestingness for Neuro-fuzzy Systems .....	242
<i>T. Wittmann, J. Ruhland, M. Eichholz</i>	
Unsupervised Profiling for Identifying Superimposed Fraud .....	251
<i>U. Murad, G. Pinkas</i>	
OPTICS-OF: Identifying Local Outliers .....	262
<i>M.M. Breunig, H.-P. Kriegel, R.T. Ng, J. Sander</i>	

## Posters

Selective Propositionalization for Relational Learning .....	271
<i>É. Alphonse, C. Rouveirol</i>	
Circle Graphs: New Visualization Tools for Text-Mining .....	277
<i>Y. Aumann, R. Feldman, Y.B. Yehuda, D. Landau, O. Liphstat, Y. Schler</i>	
On the Consistency of Information Filters for Lazy Learning Algorithms ..	283
<i>H. Brighton, C. Mellish</i>	
Using Genetic Algorithms to Evolve a Rule Hierarchy .....	289
<i>R. Cattral, F. Oppacher, D. Deugo</i>	
Mining Temporal Features in Association Rules .....	295
<i>X. Chen, I. Petrounias</i>	
The Improvement of Response Modeling: Combining Rule-Induction and Case-Based Reasoning .....	301
<i>F. Coenen, G. Swinnen, K. Vanhoof, G. Wets</i>	
Analyzing an Email Collection Using Formal Concept Analysis .....	309
<i>R. Cole, P. Eklund</i>	
Business Focused Evaluation Methods: A Case Study .....	316
<i>P. Datta</i>	
Combining Data and Knowledge by MaxEnt-Optimization of Probability Distributions .....	323
<i>W. Ertel, M. Schramm</i>	
Handling Missing Data in Trees: Surrogate Splits or Statistical Imputation? .....	329
<i>A. Feelders</i>	

Rough Dependencies as a Particular Case of Correlation: Application to the Calculation of Approximative Reducts .....	335
<i>M.C. Fernandez-Baizán, E. Menasalvas Ruiz, J.M. Peña Sánchez, S. Millán, E. Mesa</i>	
A Fuzzy Beam-Search Rule Induction Algorithm .....	341
<i>C.S. Fertig, A.A. Freitas, L.V.R. Arruda, C. Kaestner</i>	
An Innovative GA-Based Decision Tree Classifier in Large Scale Data Mining .....	348
<i>Z. Fu</i>	
Extension to C-means Algorithm for the Use of Similarity Functions.....	354
<i>J.R. García-Serrano, J.F. Martínez-Trinidad</i>	
Predicting Chemical Carcinogenesis Using Structural Information Only ...	360
<i>C.J. Kennedy, C. Giraud-Carrier, D.W. Bristol</i>	
LA - A Clustering Algorithm with an Automated Selection of Attributes, which Is Invariant to Functional Transformations of Coordinates .....	366
<i>M.V. Kiselev, S.M. Ananyan, S.B. Arseniev</i>	
Association Rule Selection in a Data Mining Environment.....	372
<i>M. Klemettinen, H. Mannila, A.I. Verkamo</i>	
Multi-relational Decision Tree Induction .....	378
<i>A.J. Knobbe, A. Siebes, D. van der Wallen</i>	
Learning of Simple Conceptual Graphs from Positive and Negative Examples .....	384
<i>S.O. Kuznetsov</i>	
An Evolutionary Algorithm Using Multivariate Discretization for Decision Rule Induction .....	392
<i>W. Kwedlo, M. Krętowski</i>	
ZigZag, a New Clustering Algorithm to Analyze Categorical Variable Cross-Classification Tables .....	398
<i>S. Lallich</i>	
Efficient Mining of High Confidence Association Rules without Support Thresholds .....	406
<i>J. Li, X. Zhang, G. Dong, K. Ramamohanarao, Q. Sun</i>	
A Logical Approach to Fuzzy Data Analysis .....	412
<i>C.-J. Liau, D.-R. Liu</i>	
AST: Support for Algorithm Selection with a CBR Approach .....	418
<i>G. Lindner, R. Studer</i>	
Efficient Shared Near Neighbours Clustering of Large Metric Data Sets ...	424
<i>S. Lodi, L. Reami, C. Sartori</i>	

Discovery of "Interesting" Data Dependencies from a Workload of SQL Statements . . . . .	430
<i>S. Lopes, J.-M. Petit, F. Toumani</i>	
Learning from Highly Structured Data by Decomposition . . . . .	436
<i>R. Mac Kinney-Romero, C. Giraud-Carrier</i>	
Combinatorial Approach for Data Binarization . . . . .	442
<i>E. Mayoraz, M. Moreira</i>	
Extending Attribute-Oriented Induction as a Key-Preserving Data Mining Method . . . . .	448
<i>M.K. Muyebe, J.A. Keane</i>	
Automated Discovery of Polynomials by Inductive Genetic Programming .	456
<i>N. Nikolaev, H. Iba</i>	
Diagnosing Acute Appendicitis with Very Simple Classification Rules . . . .	462
<i>A. Øhrn, J. Komorowski</i>	
Rule Induction in Cascade Model Based on Sum of Squares Decomposition	468
<i>T. Okada</i>	
Maintenance of Discovered Knowledge . . . . .	476
<i>M. Pěchouček, O. Štěpánková, P. Mikšovský</i>	
A Divisive Initialization Method for Clustering Algorithms . . . . .	484
<i>C. Pizzuti, D. Talia, G. Vonella</i>	
A Comparison of Model Selection Procedures for Predicting Turning Points in Financial Time Series . . . . .	492
<i>T. Poddig, K. Huber</i>	
Mining Lemma Disambiguation Rules from Czech Corpora . . . . .	498
<i>L. Popelínský, T. Pavelek</i>	
Adding Temporal Semantics to Association Rules . . . . .	504
<i>C.P. Rainsford, J.F. Roddick</i>	
Studying the Behavior of Generalized Entropy in Induction Trees Using a M-of-N Concept . . . . .	510
<i>R. Rakotomalala, S. Lallich, S. Di Palma</i>	
Discovering Rules in Information Trees . . . . .	518
<i>Z.W. Ras</i>	
Mining Text Archives: Creating Readable Maps to Structure and Describe Document Collections . . . . .	524
<i>A. Rauber, D. Merkl</i>	

XIV Table of Contents

Neuro-fuzzy Data Mining for Target Group Selection in Retail Banking . . .	530
<i>J. Ruhland, T. Wittmann</i>	
Mining Possibilistic Set-Valued Rules by Generating Prime Disjunctions . .	536
<i>A.A. Savinov</i>	
Towards Discovery of Information Granules . . . . .	542
<i>A. Skowron, J. Stepaniuk</i>	
Classification Algorithms Based on Linear Combinations of Features . . . . .	548
<i>D. Ślęzak, J. Wróblewski</i>	
Managing Interesting Rules in Sequence Mining . . . . .	554
<i>M. Spiliopoulou</i>	
Support Vector Machines for Knowledge Discovery . . . . .	561
<i>S. Sugaya, E. Suzuki, S. Tsumoto</i>	
Regression by Feature Projections . . . . .	568
<i>İ. Uysal, H.A. Güvenir</i>	
Generating Linguistic Fuzzy Rules for Pattern Classification with Genetic Algorithms . . . . .	574
<i>N. Xiong, L. Litz</i>	

**Tutorials**

Data Mining for Robust Business Intelligence Solutions . . . . .	580
<i>J. Mrazek</i>	
Query Languages for Knowledge Discovery in Databases . . . . .	582
<i>J.-F. Boulicaut</i>	
The ESPRIT Project CreditMine and Its Relevance for the Internet Market . . . . .	584
<i>S. Köhler, M. Krieger</i>	
Logics and Statistics for Association Rules and Beyond . . . . .	586
<i>P. Hájek, J. Rauch</i>	
Data Mining for the Web . . . . .	588
<i>M. Spiliopoulou</i>	
Relational Learning and Inductive Logic Programming Made Easy . . . . .	590
<i>L. De Raedt, H. Blockeel</i>	
<b>Author Index</b> . . . . .	591

Principles of Data Mining and Knowledge Discovery  
Third European Conference, PKDD'99 Prague, Czech  
Republic, September 15-18, 1999 Proceedings  
Zytkow, J.; Rauch, J. (Eds.)  
1999, XIV, 593 p., Softcover  
ISBN: 978-3-540-66490-1