

# Table of Contents

## Contributed Papers

Optimized Substructure Discovery for Semi-structured Data .....	1
<i>Kenji Abe, Shinji Kawasoe, Tatsuya Asai, Hiroki Arimura, and Setsuo Arikawa</i>	
Fast Outlier Detection in High Dimensional Spaces .....	15
<i>Fabrizio Angiulli and Clara Pizzuti</i>	
Data Mining in Schizophrenia Research – Preliminary Analysis .....	27
<i>Stefan Arnborg, Ingrid Agartz, Håkan Hall, Erik Jönsson, Anna Sillén, and Göran Sedvall</i>	
Fast Algorithms for Mining Emerging Patterns .....	39
<i>James Bailey, Thomas Manoukian, and Kotagiri Ramamohanarao</i>	
On the Discovery of Weak Periodicities in Large Time Series .....	51
<i>Christos Berberidis, Ioannis Vlahavas, Walid G. Aref, Mikhail Atallah, and Ahmed K. Elmagarmid</i>	
The Need for Low Bias Algorithms in Classification Learning from Large Data Sets .....	62
<i>Damien Brain and Geoffrey I. Webb</i>	
Mining All Non-derivable Frequent Itemsets .....	74
<i>Toon Calders and Bart Goethals</i>	
Iterative Data Squashing for Boosting Based on a Distribution-Sensitive Distance .....	86
<i>Yuta Choki and Einoshin Suzuki</i>	
Finding Association Rules with Some Very Frequent Attributes .....	99
<i>Frans Coenen and Paul Leng</i>	
Unsupervised Learning: Self-aggregation in Scaled Principal Component Space .....	112
<i>Chris Ding, Xiaofeng He, Hongyuan Zha, and Horst Simon</i>	
A Classification Approach for Prediction of Target Events in Temporal Sequences .....	125
<i>Carlotta Domeniconi, Chang-shing Perng, Ricardo Vilalta, and Sheng Ma</i>	
Privacy-Oriented Data Mining by Proof Checking .....	138
<i>Amy Felty and Stan Matwin</i>	

Choose Your Words Carefully: An Empirical Study of Feature Selection Metrics for Text Classification .....	150
<i>George Forman</i>	
Generating Actionable Knowledge by Expert-Guided Subgroup Discovery .....	163
<i>Dragan Gamberger and Nada Lavrač</i>	
Clustering Transactional Data .....	175
<i>Fosca Giannotti, Cristian Gozzi, and Giuseppe Manco</i>	
Multiscale Comparison of Temporal Patterns in Time-Series Medical Databases .....	188
<i>Shoji Hirano and Shusaku Tsumoto</i>	
Association Rules for Expressing Gradual Dependencies .....	200
<i>Eyke Hüllermeier</i>	
Support Approximations Using Bonferroni-Type Inequalities .....	212
<i>Szymon Jaroszewicz and Dan A. Simovici</i>	
Using Condensed Representations for Interactive Association Rule Mining .....	225
<i>Baptiste Jeudy and Jean-François Boulıcaut</i>	
Predicting Rare Classes: Comparing Two-Phase Rule Induction to Cost-Sensitive Boosting .....	237
<i>Mahesh V. Joshi, Ramesh C. Agarwal, and Vipin Kumar</i>	
Dependency Detection in MobiMine and Random Matrices .....	250
<i>Hillol Kargupta, Krishnamoorthy Sivakumar, and Samiran Ghosh</i>	
Long-Term Learning for Web Search Engines .....	263
<i>Charles Kemp and Kotagiri Ramamohanarao</i>	
Spatial Subgroup Mining Integrated in an Object-Relational Spatial Database .....	275
<i>Willi Klösgen and Michael May</i>	
Involving Aggregate Functions in Multi-relational Search .....	287
<i>Arno J. Knobbe, Arno Siebes, and Bart Marseille</i>	
Information Extraction in Structured Documents Using Tree Automata Induction .....	299
<i>Raymond Kosala, Jan Van den Bussche, Maurice Bruynooghe, and Hendrik Blockeel</i>	
Algebraic Techniques for Analysis of Large Discrete-Valued Datasets .....	311
<i>Mehmet Koyutürk, Ananth Grama, and Naren Ramakrishnan</i>	
Geography of Differences between Two Classes of Data .....	325
<i>Jinyan Li and Limsoon Wong</i>	

Rule Induction for Classification of Gene Expression Array Data .....	338
<i>Per Lidén, Lars Asker, and Henrik Boström</i>	
Clustering Ontology-Based Metadata in the Semantic Web .....	348
<i>Alexander Maedche and Valentin Zacharias</i>	
Iteratively Selecting Feature Subsets for Mining from High-Dimensional Databases .....	361
<i>Hiroshi Mamitsuka</i>	
SVM Classification Using Sequences of Phonemes and Syllables .....	373
<i>Gerhard Paaß, Edda Leopold, Martha Larson, Jörg Kindermann, and Stefan Eickeler</i>	
A Novel Web Text Mining Method Using the Discrete Cosine Transform .....	385
<i>Lawrence A.F. Park, Marimuthu Palaniswami, and Kotagiri Ramamohanarao</i>	
A Scalable Constant-Memory Sampling Algorithm for Pattern Discovery in Large Databases .....	397
<i>Tobias Scheffler and Stefan Wrobel</i>	
Answering the Most Correlated $N$ Association Rules Efficiently .....	410
<i>Jun Sese and Shinichi Morishita</i>	
Mining Hierarchical Decision Rules from Clinical Databases Using Rough Sets and Medical Diagnostic Model .....	423
<i>Shusaku Tsumoto</i>	
Efficiently Mining Approximate Models of Associations in Evolving Databases .....	435
<i>Adriano Veloso, Bruno Gusmão, Wagner Meira Jr., Marcio Carvalho, Srinu Parthasarathy, and Mohammed Zaki</i>	
Explaining Predictions from a Neural Network Ensemble One at a Time .....	449
<i>Robert Wall, Pádraig Cunningham, and Paul Walsh</i>	
Structuring Domain-Specific Text Archives by Deriving a Probabilistic XML DTD .....	461
<i>Karsten Winkler and Myra Spiliopoulou</i>	
Separability Index in Supervised Learning .....	475
<i>Djamel A. Zighed, Stéphane Lallich, and Fabrice Muhlenbach</i>	

## Invited Papers

Finding Hidden Factors Using Independent Component Analysis .....	488
<i>Erkki Oja</i>	

XIV     Table of Contents

Reasoning with Classifiers ..... 489  
*Dan Roth*

A Kernel Approach for Learning from Almost Orthogonal Patterns ..... 494  
*Bernhard Schölkopf, Jason Weston, Eleazar Eskin, Christina Leslie,  
and William Stafford Noble*

Learning with Mixture Models: Concepts and Applications ..... 512  
*Padhraic Smyth*

**Author Index** ..... 513

Principles of Data Mining and Knowledge Discovery  
6th European Conference, PKDD 2002, Helsinki,  
Finland, August 19-23, 2002, Proceedings  
Elomaa, T.; Mannila, H.; Toivonen, H. (Eds.)  
2002, XIV, 514 p., Softcover  
ISBN: 978-3-540-44037-6