

# Table of Contents

## Invited Papers

Random Matrices in Data Analysis . . . . .	1
<i>Dimitris Achlioptas</i>	
Data Privacy . . . . .	8
<i>Rakesh Agrawal</i>	
Breaking Through the Syntax Barrier: Searching with Entities and Relations . . . . .	9
<i>Soumen Chakrabarti</i>	
Real-World Learning with Markov Logic Networks . . . . .	17
<i>Pedro Domingos</i>	
Strength in Diversity: The Advance of Data Analysis . . . . .	18
<i>David J. Hand</i>	

## Contributed Papers

Mining Positive and Negative Association Rules: An Approach for Confined Rules . . . . .	27
<i>Maria-Luiza Antonie and Osmar R. Zaïane</i>	
An Experiment on Knowledge Discovery in Chemical Databases . . . . .	39
<i>Sandra Berasaluze, Claude Laurenço, Amedeo Napoli, and Gilles Niel</i>	
Shape and Size Regularization in Expectation Maximization and Fuzzy Clustering . . . . .	52
<i>Christian Borgelt and Rudolf Kruse</i>	
Combining Multiple Clustering Systems . . . . .	63
<i>Constantinos Boulis and Mari Ostendorf</i>	
Reducing Data Stream Sliding Windows by Cyclic Tree-Like Histograms . .	75
<i>Francesco Buccafurri and Gianluca Lax</i>	
A Framework for Data Mining Pattern Management . . . . .	87
<i>Barbara Catania, Anna Maddalena, Maurizio Mazza, Elisa Bertino, and Stefano Rizzi</i>	
Spatial Associative Classification at Different Levels of Granularity: A Probabilistic Approach . . . . .	99
<i>Michelangelo Ceci, Annalisa Appice, and Donato Malerba</i>	

AutoPart: Parameter-Free Graph Partitioning and Outlier Detection . . . . .	112
<i>Deepayan Chakrabarti</i>	
Properties and Benefits of Calibrated Classifiers . . . . .	125
<i>Ira Cohen and Moises Goldszmidt</i>	
A Tree-Based Approach to Clustering XML Documents by Structure . . . . .	137
<i>Gianni Costa, Giuseppe Manco, Riccardo Ortale, and Andrea Tagarelli</i>	
Discovery of Regulatory Connections in Microarray Data . . . . .	149
<i>Michael Egmont-Petersen, Wim de Jonge, and Arno Siebes</i>	
Learning from Little: Comparison of Classifiers Given Little Training . . . . .	161
<i>George Forman and Ira Cohen</i>	
Geometric and Combinatorial Tiles in 0–1 Data . . . . .	173
<i>Aristides Gionis, Heikki Mannila, and Jouni K. Seppänen</i>	
Document Classification Through Interactive Supervision of Document and Term Labels . . . . .	185
<i>Shantanu Godbole, Abhay Harpale, Sunita Sarawagi, and Soumen Chakrabarti</i>	
Classifying Protein Fingerprints . . . . .	197
<i>Melanie Hilario, Alex Mitchell, Jee-Hyub Kim, Paul Bradley, and Terri Attwood</i>	
Finding Interesting Pass Patterns from Soccer Game Records . . . . .	209
<i>Shoji Hirano and Shusaku Tsumoto</i>	
Discovering Unexpected Information for Technology Watch . . . . .	219
<i>François Jacquenet and Christine Largeron</i>	
Scalable Density-Based Distributed Clustering . . . . .	231
<i>Eshref Januzaj, Hans-Peter Kriegel, and Martin Pfeifle</i>	
Summarization of Dynamic Content in Web Collections . . . . .	245
<i>Adam Jatowt and Mitsuru Ishizuka</i>	
Mining Thick Skylines over Large Databases . . . . .	255
<i>Wen Jin, Jiawei Han, and Martin Ester</i>	
Ensemble Feature Ranking . . . . .	267
<i>Kees Jong, Jérémie Mary, Antoine Cornuéjols, Elena Marchiori, and Michèle Sebag</i>	
Privately Computing a Distributed $k$ -nn Classifier . . . . .	279
<i>Murat Kantarcioğlu and Chris Clifton</i>	
Incremental Nonlinear PCA for Classification . . . . .	291
<i>Byung Joo Kim and Il Kon Kim</i>	

A Spectroscopy of Texts for Effective Clustering . . . . .	301
<i>Wenyuan Li, Wee-Keong Ng, Kok-Leong Ong, and Ee-Peng Lim</i>	
Constraint-Based Mining of Episode Rules and Optimal Window Sizes . . . .	313
<i>Nicolas Méger and Christophe Rigotti</i>	
Analysing Customer Churn in Insurance Data – A Case Study . . . . .	325
<i>Katharina Morik and Hanna Köpcke</i>	
Nomograms for Visualization of Naive Bayesian Classifier . . . . .	337
<i>Martin Možina, Janez Demšar, Michael Kattan, and Blaž Zupan</i>	
Using a Hash-Based Method for Apriori-Based Graph Mining . . . . .	349
<i>Phu Chien Nguyen, Takashi Washio, Kouzou Ohara, and Hiroshi Motoda</i>	
Evaluation of Rule Interestingness Measures with a Clinical Dataset on Hepatitis . . . . .	362
<i>Miho Ohsaki, Shinya Kitaguchi, Kazuya Okamoto, Hideto Yokoi, and Takahira Yamaguchi</i>	
Classification in Geographical Information Systems . . . . .	374
<i>Salvatore Rinzivillo and Franco Turini</i>	
Digging into Acceptor Splice Site Prediction: An Iterative Feature Selection Approach . . . . .	386
<i>Yvan Saeys, Sven Degroeve, and Yves Van de Peer</i>	
Itemset Classified Clustering . . . . .	398
<i>Jun Sese and Shinichi Morishita</i>	
Combining Winnow and Orthogonal Sparse Bigrams for Incremental Spam Filtering . . . . .	410
<i>Christian Siefkes, Fidelis Assis, Shalendra Chhabra, and William S. Yerazunis</i>	
Asynchronous and Anticipatory Filter-Stream Based Parallel Algorithm for Frequent Itemset Mining . . . . .	422
<i>Adriano Veloso, Wagner Meira Jr., Renato Ferreira, Dorgival Guedes Neto, and Srinivasan Parthasarathy</i>	
A Quantification of Cluster Novelty with an Application to Martian Topography . . . . .	434
<i>Ricardo Vilalta, Tom Stepinski, Muralikrishna Achari, and Francisco Ocegueda-Hernandez</i>	
Density-Based Spatial Clustering in the Presence of Obstacles and Facilitators . . . . .	446
<i>Xin Wang, Camilo Rostoker, and Howard J. Hamilton</i>	

Text Mining for Finding Functional Community of Related Genes  
Using TCM Knowledge ..... 459  
*Zhaohui Wu, Xuezhong Zhou, Baoyan Liu, and Junli Chen*

Dealing with Predictive-but-Unpredictable Attributes  
in Noisy Data Sources ..... 471  
*Ying Yang, Xindong Wu, and Xingquan Zhu*

A New Scheme on Privacy Preserving Association Rule Mining ..... 484  
*Nan Zhang, Shengquan Wang, and Wei Zhao*

**Posters**

A Unified and Flexible Framework for Comparing Simple  
and Complex Patterns ..... 496  
*Ilaria Bartolini, Paolo Ciaccia, Irene Ntoutsis, Marco Patella,  
and Yannis Theodoridis*

Constructing (Almost) Phylogenetic Trees  
from Developmental Sequences Data ..... 500  
*Ronnie Bathoorn and Arno Siebes*

Learning from Multi-source Data ..... 503  
*Elisa Fromont, Marie-Odile Cordier, and René Quiniou*

The Anatomy of SnakeT: A Hierarchical Clustering Engine  
for Web-Page Snippets ..... 506  
*Paolo Ferragina and Antonio Gulli*

COCOA: Compressed Continuity Analysis for Temporal Databases ..... 509  
*Kuo-Yu Huang, Chia-Hui Chang, and Kuo-Zui Lin*

Discovering Interpretable Muscle Activation Patterns  
with the Temporal Data Mining Method ..... 512  
*Fabian Mörchen, Alfred Ultsch, and Olaf Hoos*

A Tolerance Rough Set Approach to Clustering Web Search Results ..... 515  
*Chi Lang Ngo and Hung Son Nguyen*

Improving the Performance of the RISE Algorithm ..... 518  
*Aloisio Carlos de Pina and Gerson Zaverucha*

Mining History of Changes to Web Access Patterns ..... 521  
*Qiankun Zhao and Sourav S. Bhowmick*

**Demonstration Papers**

Visual Mining of Spatial Time Series Data ..... 524  
*Gennady Andrienko, Natalia Andrienko, and Peter Gatalsky*

Detecting Driving Awareness .....	528
<i>Bruno Apolloni, Andrea Brega, Dario Malchiodi, and Cristian Mesiano</i>	
An Effective Recommender System for Highly Dynamic and Large Web Sites .....	531
<i>Ranieri Baraglia, Francesco Merlo, and Fabrizio Silvestri</i>	
SemanticTalk: Software for Visualizing Brainstorming Sessions and Thematic Concept Trails on Document Collections .....	534
<i>Chris Biemann,, Karsten Böhm, Gerhard Heyer, and Ronny Melz</i>	
Orange: From Experimental Machine Learning to Interactive Data Mining .....	537
<i>Janez Demšar, Blaž Zupan, Gregor Leban, and Tomaz Curk</i>	
Terrorist Detection System .....	540
<i>Yuval Elovici, Abraham Kandel, Mark Last, Bracha Shapira, Omer Zaafrany, Moti Schneider, and Menahem Friedman</i>	
Experimenting SnakeT: A Hierarchical Clustering Engine for Web-Page Snippets.....	543
<i>Paolo Ferragina and Antonio Gulli</i>	
HIClass: Hyper-interactive Text Classification by Interactive Supervision of Document and Term Labels .....	546
<i>Shantanu Godbole, Abhay Harpale, Sunita Sarawagi, and Soumen Chakrabarti</i>	
BALIOS – The Engine for Bayesian Logic Programs.....	549
<i>Kristian Kersting and Uwe Dick</i>	
SEWeP: A Web Mining System Supporting Semantic Personalization .....	552
<i>Stratos Paulakis, Charalampos Lampos, Magdalini Eirinaki, and Michalis Vazirgiannis</i>	
SPIN! Data Mining System Based on Component Architecture.....	555
<i>Alexandr Savinov</i>	
<b>Author Index .....</b>	<b>559</b>

Knowledge Discovery in Databases: PKDD 2004  
8th European Conference on Principles and Practice of  
Knowledge Discovery in Databases, Pisa, Italy,  
September 20-24, 2004, Proceedings  
Boulıcaut, J.-F.; Esposito, F.; Giannotti, F.; Pedreschi, D.  
(Eds.)  
2004, XIX, 562 p., Softcover  
ISBN: 978-3-540-23108-0