The seventh event of the International Conference on Machine Learning and Data Mining (MLDM) was held in New York (www.mldm.de) running under the umbrella of the world congress "The Frontiers in Intelligent Data and Signal Analysis, DSA2011.”

For this edition the Program Committee received 170 submissions. After the peer-review process, we accepted 56 high-quality papers for oral presentation, and from these 44 are included in this proceedings book. The topics range from theoretical topics for classification, clustering, association rule and pattern mining to specific data mining methods for the different multimedia data types such as image mining, text mining, video mining and Web mining. Extended versions of selected papers will appear in the international journal Transactions on Machine Learning and Data Mining (www.ibai-publishing.org/journal/mldm).

Fourteen papers were selected for poster presentation and are published in the MLDM Poster Proceedings by ibai-publishing (www.ibai-publishing.org).

A tutorial on Data Mining and a tutorial on Case-Based Reasoning were held before the conference.

We were pleased to give out the best paper award for MLDM for the third time this year. The final decision was made by the Best Paper Award Committee based on the presentation by the authors and the discussion with the auditorium. The ceremony took place at the end of the conference. This prize is sponsored by ibai solutions (www.ibai-solutions.de), one of the leading companies in data mining for marketing, Web mining and e-commerce.

The conference was rounded up by an outlook of new challenging topics in machine learning and data mining before the Best Paper Award Ceremony.

We thank the members of the Institute of Applied Computer Sciences, Leipzig, Germany (www.ibai-institut.de) who handled the conference as secretariat. We appreciate the help and understanding of the editorial staff at Springer, and in particular Alfred Hofmann, who supported the publication of these proceedings in the LNAI series.

Last, but not least, we wish to thank all the speakers and participants who contributed to the success of the conference. St. Petersburg will host the next world congress on “The Frontiers in Intelligent Data and Signal Analysis, DSA2012” (www.worldcongressdsa.com) in 2012, combining under its roof the following events: International Conference on Machine Learning and Data Mining (MLDM); the Industrial Conference on Data Mining (ICDM), and the International Conference on Mass Data Analysis of Signals and Images in Medicine, Biotechnology, Chemistry and Food Industry (MDA).

August 2011

Petra Perner
Organization

Chair

Petra Perner  
IBAI Leipzig, Germany

Program Committee

Agnar Aamodt  
NTNU, Norway
Jacky Baltes  
University of Manitoba, Canada
Max Bramer  
University of Portsmouth, UK
Horst Bunke  
University of Bern, Switzerland
Krzysztof Cios  
University of Colorado, USA
Horst Bunke  
University of Bern, Switzerland
Krzysztof Cios  
University of Colorado, USA
John Debenham  
University of Technology, Australia
Christoph F. Eick  
University of Houston, USA
Ana Fred  
Technical University of Lisbon, Portugal
Giorgio Giacinto  
University of Cagliari, Italy
Makato Haraguchi  
Hokkaido University Sapporo, Japan
Robert J. Hilderman  
University of Regina, Canada
Tin Kam Ho  
Bell Laboratories, USA
Atsushi Imiya  
Chiba University, Japan
Horace Ip  
City University, Hong Kong
Abraham Kandel  
University of South Florida, USA
Dimitrios A. Karras  
Chalkis Institute of Technology, Greece
Adam Krzyzak  
Concordia University, Montreal, Canada
Lukasz Kurgan  
University of Alberta, Canada
Longin Jan Latecki  
Temple University Philadelphia, USA
Tao Li  
Florida International University, USA
Brian Lovell  
University of Queensland, Australia
Mariofanna Milanova  
University of Arkansas at Little Rock, USA
Thang V. Pham  
University of Amsterdam, The Netherlands
Maria da Graca Pimentel  
Universidade de Sao Paulo, Brazil
Petia Radeva  
Universitat Autonoma de Barcelona, Spain
Michael Richter  
University of Calgary, Canada
Fabio Roli  
University of Cagliari, Italy
Linda Shapiro  
University of Washington, USA
Sameer Singh  
Loughborough University, UK
David Steinberg  
Tel Aviv University, Israel
Francesco Tortorella  
Università degli Studi di Cassino, Italy
Patrick Wang  
Northeastern University, USA
### Additional Reviewers

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richard Blake</td>
<td>NTNU, Norway</td>
</tr>
<tr>
<td>Gleb Sizov</td>
<td>NTNU, Norway</td>
</tr>
<tr>
<td>Pinar Özturk</td>
<td>NTNU, Norway</td>
</tr>
<tr>
<td>Pål Sætrom (Paal Saetrom)</td>
<td>NTNU, Norway</td>
</tr>
<tr>
<td>Helge Langseth</td>
<td>NTNU, Norway</td>
</tr>
<tr>
<td>Tore Bruland</td>
<td>NTNU, Norway</td>
</tr>
<tr>
<td>Ajita Rattani</td>
<td>University of Cagliari, Italy</td>
</tr>
<tr>
<td>Ignazio Pillai</td>
<td>University of Cagliari, Italy</td>
</tr>
<tr>
<td>Davide Ariu</td>
<td>University of Cagliari, Italy</td>
</tr>
<tr>
<td>Luca Didaci</td>
<td>University of Cagliari, Italy</td>
</tr>
<tr>
<td>Roberto Tronci</td>
<td>University of Cagliari, Italy</td>
</tr>
<tr>
<td>Battista Biggio</td>
<td>University of Cagliari, Italy</td>
</tr>
</tbody>
</table>
# Table of Contents

## Classification and Decision Theory

Quadratically Constrained Maximum a Posteriori Estimation for Binary Classifier .................................................. 1  
*Tatsuya Yokota and Yukihiko Yamashita*

Hubness-Based Fuzzy Measures for High-Dimensional $k$-Nearest Neighbor Classification ........................................... 16  
*Nenad Tomašev, Miloš Radovanović, Dunja Mladenić, and Mirjana Ivanović*

Decisions: Algebra and Implementation ............................................. 31  
*Antonina Danylenko, Jonas Lundberg, and Welf Löwe*

Smoothing Multinomial Naïve Bayes in the Presence of Imbalance....... 46  
*Alexander Y. Liu and Cheryl E. Martin*

ACE-Cost: Acquisition Cost Efficient Classifier by Hybrid Decision Tree with Local SVM Leaves ................................. 60  
*Liyun Li, Umut Topkara, and Nasir Memon*

Informative Variables Selection for Multi-relational Supervised Learning .......................................................... 75  
*Dhafer Lahbib, Marc Boullé, and Dominique Laurent*

Separability of Split Value Criterion with Weighted Separation Gains . . 88  
*Krzysztof Grąbczewski*

Granular Instances Selection for Fuzzy Modeling ............................. 99  
*S. Sakinah S. Ahmad and Witold Pedrycz*

Parameter-Free Anomaly Detection for Categorical Data .................. 112  
*Shu Wu and Shengrui Wang*

Fuzzy Semi-supervised Support Vector Machines ............................ 127  
*Houda Benbrahim*

GENCCS: A Correlated Group Difference Approach to Contrast Set Mining ........................................................................ 140  
*Mondelle Simeon and Robert Hilderman*

Collective Classification Using Heterogeneous Classifiers ................. 155  
*Zehra Cataltepe, Abdullah Sonmez, Kadriye Baglioglu, and Ayse Erzan*
Spherical Nearest Neighbor Classification: Application to Hyperspectral Data ........................................................... 170
   Dalton Lunga and Okan Ersoy

Adaptive Kernel Diverse Density Estimate for Multiple Instance Learning ........................................................... 185
   Tao Xu, Iker Gondra, and David Chiu

Boosting Inspired Process for Improving AUC ......................... 199
   Victor S. Sheng and Rahul Tada

Theory of Learning

Investigation in Transfer Learning: Better Way to Apply Transfer Learning between Agents ........................................... 210
   Luiz Antonio Celiberto Junior and Jackson P. Matsuura

Exploration Strategies for Learned Probabilities in Smart Terrain ...... 224
   John Sullins

Sensitivity Analysis for Weak Constraint Generation .................. 239
   Jamshaid G. Mohebzada, Michael M. Richter, and Guenther Ruhe

Dictionary Learning Based on Laplacian Score in Sparse Coding ...... 253
   Jin Xu and Hong Man

Clustering

A Practical Approach for Clustering Transaction Data .................. 265
   Mohamed Bouguessa

Hierarchical Clustering with High Order Dissimilarities ............... 280
   Helena Aidos and Ana Fred

Clust-XPaths: Clustering of XML Paths .................................. 294
   Amina Madani, Omar Boussaid, and Djamel Eddine Zegour

Comparing Clustering and Metaclustering Algorithms ................... 306
   Elio Lozano and Edgar Acuña

Applications in Medicine

Detection of Phenotypes in Microarray Data Using Force-Directed Placement Transforms ........................................... 320
   Dragana Veljkovic Perez and Kay A. Robbins

On the Temporal Behavior of EEG Recorded during Real Finger Movement ........................................... 335
   Bashar Awwad Shiek Hasan
A Machine Learning and Data Mining Framework to Enable Evolutionary Improvement in Trauma Triage ....................... 348
   Douglas A. Talbert, Matt Honeycutt, and Steve Talbert

A Decision Support System Based on the Semantic Analysis of Melanoma Images Using Multi-elitist PSO and SVM .................. 362
   Weronika Piątkowska, Jerzy Martyna, Leszek Nowak, and Karol Przystalski

WebMining/Information Mining

Authorship Similarity Detection from Email Messages .................... 375
   Xiaoling Chen, Peng Hao, R. Chandramouli, and K.P. Subbalakshmi

An Investigation Concerning the Generation of Text Summarisation Classifiers Using Secondary Data ............................... 387
   Matias Garcia-Constantino, Frans Coenen, P.-J. Noble, Alan Radford, Christian Setzkorn, and Aine Tierney

Comparing the One-Vs-One and One-Vs-All Methods in Benthic Macroinvertebrate Image Classification .............................. 399
   Henry Joutsijoki and Martti Juhola

Incremental Web-Site Boundary Detection Using Random Walks .......... 414
   Ayesh Alshukri, Frans Coenen, and Michele Zito

Discovering Text Patterns by a New Graphic Model ........................ 428
   Minhua Huang and Robert M. Haralick

Topic Sentiment Change Analysis ........................................... 443
   Yu Jiang, Weiyi Meng, and Clement Yu

Adaptive Context Modeling for Deception Detection in Emails.......... 458
   Peng Hao, Xiaoling Chen, Na Cheng, R. Chandramouli, and K.P. Subbalakshmi

Contrasting Correlations by an Efficient Double-Clique Condition ...... 469
   Aixiang Li, Makoto Haraguchi, and Yoshiaki Okubo

Machine Learning and Image Mining

Estimating Image Segmentation Difficulty ................................ 484
   Dingding Liu, Yingen Xiong, Kari Pulli, and Linda Shapiro

Mining Spatial Trajectories Using Non-parametric Density Functions ... 496
   Chun-Sheng Chen, Christoph F. Eick, and Nouhad J. Rizk
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploring Synergetic Effects of Dimensionality Reduction and Resampling Tools on Hyperspectral Imagery Data Classification</td>
<td>511</td>
</tr>
<tr>
<td>J.S. Sánchez, V. García, and R.A. Mollineda</td>
<td></td>
</tr>
<tr>
<td>A Comparison between Haralick’s Texture Descriptor and the Texture Descriptor Based on Random Sets for Biological Images</td>
<td>524</td>
</tr>
<tr>
<td>Anja Attig and Petra Perner</td>
<td></td>
</tr>
<tr>
<td><strong>Time Series and Frequent Item Set Mining</strong></td>
<td></td>
</tr>
<tr>
<td>Unsupervised Discovery of Motifs under Amplitude Scaling and Shifting in Time Series Databases</td>
<td>539</td>
</tr>
<tr>
<td>Tom Armstrong and Eric Drewniak</td>
<td></td>
</tr>
<tr>
<td>Static Load Balancing of Parallel Mining of Frequent Itemsets Using Reservoir Sampling</td>
<td>553</td>
</tr>
<tr>
<td>Robert Kessl</td>
<td></td>
</tr>
<tr>
<td>GA-TVRC: A Novel Relational Time Varying Classifier to Extract Temporal Information Using Genetic Algorithms</td>
<td>568</td>
</tr>
<tr>
<td>İsmail Güneş, Zehra Çataltepe, and Şule Gündüz Öğüdükü</td>
<td></td>
</tr>
<tr>
<td><strong>Aspects of Machine Learning and Data Mining</strong></td>
<td></td>
</tr>
<tr>
<td>Detection of Communities and Bridges in Weighted Networks</td>
<td>584</td>
</tr>
<tr>
<td>Tanwistha Saha, Carlotta Domeniconi, and Huzefa Rangwala</td>
<td></td>
</tr>
<tr>
<td>Techniques for Improving Filters in Power Grid Contingency Analysis</td>
<td>599</td>
</tr>
<tr>
<td>Robert Adolf, David Haglin, Mahantesh Halappanavar, Yousu Chen, and Zhenyu Huang</td>
<td></td>
</tr>
<tr>
<td><strong>Author Index</strong></td>
<td>613</td>
</tr>
</tbody>
</table>