Petra Perner (Ed.)

Advances in Data Mining
Applications and Theoretical Aspects

10th Industrial Conference, ICDM 2010
Berlin, Germany, July 12-14, 2010
Proceedings
Preface

These are the proceedings of the tenth event of the Industrial Conference on Data Mining ICDM held in Berlin (www.data-mining-forum.de).

For this edition the Program Committee received 175 submissions. After the peer-review process, we accepted 49 high-quality papers for oral presentation that are included in this book. The topics range from theoretical aspects of data mining to applications of data mining such as on multimedia data, in marketing, finance and telecommunications, in medicine and agriculture, and in process control, industry and society. Extended versions of selected papers will appear in the international journal *Transactions on Machine Learning and Data Mining* (www.ibai-publishing.org/journal/mldm).

Ten papers were selected for poster presentations and are published in the *ICDM Poster Proceeding* Volume by ibai-publishing (www.ibai-publishing.org).

In conjunction with ICDM four workshops were held on special hot application-oriented topics in data mining: Data Mining in Marketing DMM, Data Mining in LifeScience DMLS, the Workshop on Case-Based Reasoning for Multimedia Data CBR-MD, and the Workshop on Data Mining in Agriculture DMA. The Workshop on Data Mining in Agriculture ran for the first time this year. All workshop papers will be published in the workshop proceedings by ibai-publishing (www.ibai-publishing.org). Selected papers of CBR-MD will be published in a special issue of the international journal *Transactions on Case-Based Reasoning* (www.ibai-publishing.org/journal/cbr).

We were pleased to give out the best paper award for ICDM again 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 data mining companies in data mining for marketing, Web mining and E-Commerce.

The conference was rounded up by an outlook on new challenging topics in 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. The next conference in the series will be held in 2011 in New York during the world congress “The Frontiers in Intelligent Data and Signal Analysis, DSA2011” (www.worldcongressdsa.com) that brings together the
International Conferences 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).

July 2010  

Petra Perner
Industrial Conference on Data Mining, ICDM 2010

Chair

Petra Perner
IBaI Leipzig, Germany

Program Committee

Klaus-Peter Adlassnig
Medical University of Vienna, Austria
Andrea Ahlemeyer-Stubbe
ENBIS, The Netherlands
Klaus-Dieter Althoff
University of Hildesheim, Germany
Chid Apte
IBM Yorktown Heights, USA
Eva Armengol
IIA CSIC, Spain
Bart Baesens
KU Leuven, Belgium
Isabelle Bichindaritz
University of Washington, USA
Leon Bobrowski
Bialystok Technical University, Poland
Marc Boullé
France Télécom, France
Henning Christiansen
Roskilde University, Denmark
Shirley Coleman
University of Newcastle, UK
Juan M. Corchado
Universidad de Salamanca, Spain
Antonio Dourado
University of Coimbra, Portugal
Peter Funk
Mälardalen University, Sweden
Brent Gordon
NASA Goddard Space Flight Center, USA
Gary F. Holness
Quantum Leap Innovations Inc., USA
Eyke Hüllermeier
University of Marburg, Germany
Piotr Jedrzejowicz
Gdynia Maritime University, Poland
Janusz Kacprzyk
Polish Academy of Sciences, Poland
Mehmed Kantardzic
University of Louisville, USA
Ron Kenett
KPA Ltd., Israel
Mineichi Kudo
Hokkaido University, Japan
David Manzano Macho
Ericsson Research Spain, Spain
Eduardo F. Morales
INAOE, Ciencias Computacionales, Mexico
Stefania Montani
Università del Piemonte Orientale, Italy
Jerry Oglesby
SAS Institute Inc., USA
Eric Pauwels
CWI Utrecht, The Netherlands
Mykola Pechenizkiy
Eindhoven University of Technology,
The Netherlands
Ashwin Ram
Georgia Institute of Technology, USA
Tim Rey
Dow Chemical Company, USA
Rainer Schmidt
University of Rostock, Germany
Yuval Shahar
Ben Gurion University, Israel
David Taniar
Monash University, Australia
Stijn Viaene  
KU Leuven, Belgium

Rob A. Vingerhoeds  
Ecole Nationale d'Ingénieurs de Tarbes, France

Yanbo J. Wang  
Information Management Center, China
Minsheng Banking Corporation Ltd., China

Claus Weihs  
University of Dortmund, Germany

Terry Windeatt  
University of Surrey, UK
# Table of Contents

## Invited Talk

Moving Targets: When Data Classes Depend on Subjective Judgement, or They Are Crafted by an Adversary to Mislead Pattern Analysis Algorithms - The Cases of Content Based Image Retrieval and Adversarial Classification  
*Giorgio Giacinto*

## Theoretical Aspects of Data Mining

Bootstrap Feature Selection for Ensemble Classifiers  
*Rakkrit Duangsoithong and Terry Windeatt*

Evaluating the Quality of Clustering Algorithms Using Cluster Path Lengths  
*Faraz Zaidi, Daniel Archambault, and Guy Melançon*

Finding Irregularly Shaped Clusters Based on Entropy  
*Angel Kuri-Morales and Edwin Aldana-Bobadilla*

Fuzzy Conceptual Clustering  
*Petra Perner and Anja Attig*

Mining Concept Similarities for Heterogeneous Ontologies  
*Konstantin Todorov, Peter Geibel, and Kai-Uwe Kühnberger*

Re-mining Positive and Negative Association Mining Results  
*Ayhan Demiriz, Gurdal Ertek, Tankut Atan, and Ufuk Kula*

Multi-Agent Based Clustering: Towards Generic Multi-Agent Data Mining  
*Santhana Chaimontree, Katie Atkinson, and Frans Coenen*

Describing Data with the Support Vector Shell in Distributed Environments  
*Peng Wang and Guojun Mao*

Robust Clustering Using Discriminant Analysis  
*Vasudha Bhatnagar and Sangeeta Ahuja*
New Approach in Data Stream Association Rule Mining Based on Graph Structure .......................................................... 158
   Samad Gahderi Mojaveri, Esmaeil Mirzaeian, Zarrintaj Bornaee, and Saeed Ayat

Multimedia Data Mining

Fast Training of Neural Networks for Image Compression ................. 165
   Yevgeniy Bodyanskiy, Paul Grimm, Sergey Mashtalir, and Vladimir Vinarski

Processing Handwritten Words by Intelligent Use of OCR Results ...... 174
   Benjamin Mund and Karl-Heinz Steinke

Saliency-Based Candidate Inspection Region Extraction in Tape Automated Bonding ....................................................... 186
   Martina Dümcke and Hiroki Takahashi

Image Classification Using Histograms and Time Series Analysis: A Study of Age-Related Macular Degeneration Screening in Retinal Image Data ..................................................................................... 197
   Mohd Hanafi Ahmad Hijazi, Frans Coenen, and Yalin Zheng

Entropic Quadtrees and Mining Mars Craters .................................. 210
   Rosanne Vetro and Dan A. Simovici

Hybrid DIAAF/RS: Statistical Textual Feature Selection for Language-Independent Text Classification ........................................ 222
   Yanbo J. Wang, Fan Li, Frans Coenen, Robert Sanderson, and Qin Xin

Multimedia Summarization in Law Courts: A Clustering-Based Environment for Browsing and Consulting Judicial Folders .......... 237
   E. Fersini, E. Messina, and F. Archetti

Comparison of Redundancy and Relevance Measures for Feature Selection in Tissue Classification of CT Images .......................... 248
   Benjamin Auffarth, Maite López, and Jesús Cerquides

Data Mining in Marketing

Quantile Regression Model for Impact Toughness Estimation ............. 263
   Satu Tamminen, Ilmari Juutilainen, and Juha Röning

Mining for Paths in Flow Graphs .................................................. 277
   Adam Jocksch, José Nelson Amaral, and Marcel Mitran
Combining Unsupervised and Supervised Data Mining Techniques for Conducting Customer Portfolio Analysis .......................... 292
   Zhiyuan Yao, Annika H. Holmbom, Tomas Eklund, and Barbro Back

Managing Product Life Cycle with MultiAgent Data Mining System .... 308
   Serge Parshutin

Modeling Pricing Strategies Using Game Theory and Support Vector Machines .......................................................... 323
   Cristián Bravo, Nicolás Figueroa, and Richard Weber

Data Mining in Industrial Processes

   Yuehjen E. Shao, Chi-Jie Lu, and Yu-Chiun Wang

Dynamic Pattern Extraction of Parameters in Laser Welding Process ... 350
   Gissel Velarde and Christian Binroth

Trajectory Clustering for Vibration Detection in Aircraft Engines ...... 362
   Aurélien Hazan, Michel Verleysen, Marie Cottrell, and Jérôme Lacaille

Episode Rule-Based Prognosis Applied to Complex Vacuum Pumping Systems Using Vibratory Data........................................ 376
   Florent Martin, Nicolas Mégé, Sylvie Galichet, and Nicolas Becourt

Predicting Disk Failures with HMM- and HSMM-Based Approaches .... 390
   Ying Zhao, Xiang Liu, Siqing Gan, and Weimin Zheng

Aircraft Engine Health Monitoring Using Self-Organizing Maps ....... 405
   Étienne Côme, Marie Cottrell, Michel Verleysen, and Jérôme Lacaille

Data Mining in Medicine

Finding Temporal Patterns in Noisy Longitudinal Data: A Study in Diabetic Retinopathy .......................................................... 418
   Vassiliki Somaraki, Deborah Broadbent, Frans Coenen, and Simon Harding

Selection of High Risk Patients with Ranked Models Based on the CPL Criterion Functions ................................................. 432
   Leon Bobrowski

Medical Datasets Analysis: A Constructive Induction Approach ....... 442
   Wiesław Paja and Mariusz Wrzesień
Data Mining in Agriculture

Regression Models for Spatial Data: An Example from Precision Agriculture .............................................. 450
  Georg Ruß and Rudolf Kruse

Trend Mining in Social Networks: A Study Using a Large Cattle Movement Database ........................................ 464
  Puteri N.E. Nohuddin, Rob Christley, Frans Coenen, and Christian Setzkorn

WebMining

Spam Email Filtering Using Network-Level Properties ............... 476
  Paulo Cortez, André Correia, Pedro Sousa, Miguel Rocha, and Miguel Rio

Domain-Specific Identification of Topics and Trends in the Blogosphere ..................................................... 490
  Rafael Schirru, Darko Obradović, Stephan Baumann, and Peter Wortmann

Combining Business Process and Data Discovery Techniques for Analyzing and Improving Integrated Care Pathways .............. 505
  Jonas Poelmans, Guido Dedene, Gerda Verheyden, Herman Van der Mussele, Stijn Viaene, and Edward Peters

Interest-Determining Web Browser ........................................ 518
  Khaled Bashir Shaban, Joannes Chan, and Raymond Szeto

Web-Site Boundary Detection ............................................. 529
  Ayesh Alshukri, Frans Coenen, and Michele Zito

Data Mining in Finance

An Application of Element Oriented Analysis Based Credit Scoring .... 544
  Yihao Zhang, Mehmet A. Orgun, Rohan Baxter, and Weiqiang Lin

A Semi-supervised Approach for Reject Inference in Credit Scoring Using SVMs .................................................. 558
  Sebastián Maldonado and Gonzalo Paredes

Aspects of Data Mining

Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool ............................................. 572
  Paulo Cortez
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Orange Customer Analysis Platform</td>
<td>584</td>
</tr>
<tr>
<td>Raphaël Féraud, Marc Boullé, Fabrice Clérot, Françoise Fessant, and Vincent Lemaire</td>
<td></td>
</tr>
<tr>
<td>Semi-supervised Learning for False Alarm Reduction</td>
<td>595</td>
</tr>
<tr>
<td>Chien-Yi Chiu, Yuh-Jye Lee, Chien-Chung Chang, Wen-Yang Luo, and Hsiu-Chuan Huang</td>
<td></td>
</tr>
<tr>
<td>Learning from Humanoid Cartoon Designs</td>
<td>606</td>
</tr>
<tr>
<td>Mining Relationship Associations from Knowledge about Failures Using Ontology and Inference</td>
<td>617</td>
</tr>
<tr>
<td>Weisen Guo and Steven B. Kraines</td>
<td></td>
</tr>
</tbody>
</table>

**Data Mining for Network Performance Monitoring**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Prediction in Network Monitoring Systems: Performing Sequential Pattern Mining in Osnius Monitoring Tool</td>
<td>632</td>
</tr>
<tr>
<td>Rafael García, Luis Llana, Constantino Malagón, and Jesús Pancorbo</td>
<td></td>
</tr>
<tr>
<td>Selection of Effective Network Parameters in Attacks for Intrusion Detection</td>
<td>643</td>
</tr>
<tr>
<td>Gholam Reza Zargar and Peyman Kabiri</td>
<td></td>
</tr>
</tbody>
</table>

**Author Index** | 653  |