Preface

The 11th event of the Industrial Conference on Data Mining (ICDM) was held in New York (www.data-mining-forum.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 104 submissions. After the peer-review process, we accepted 33 high-quality papers for oral presentation, and from these 24 are included in this proceedings book. The topics range from theoretical aspects of data mining to applications of data mining such as on multimedia data, in marketing, finance and telecommunication, 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).

Fourteen papers were selected for poster presentation and five for industry paper presentation, and they are published in the *ICDM Poster and Industry Proceedings by ibai-publishing* (www.ibai-publishing.org).

In conjunction with ICDM four workshops were run focusing on special hot application-oriented topics in data mining: Data Mining in Marketing (DMM), Data Mining in Life Science (DMLS), the Workshop on Case-Based Reasoning for Multimedia Data (CBR-MD), and the Workshop on Data Mining in Agriculture (DMA). All workshop papers appear in the *workshop proceedings* published 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 ICDM for the fifth 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 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. See you in 2012 to the next world congress on “The Frontiers in Intelligent Data and Signal Analysis, DSA2012” (www.worldcongressdsa.com) in 2012, combining under its roof the three
following events: the 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

Klaus-Peter Adlassnig, Medical University of Vienna, Austria
Andrea Ahlemeyer-Stubbe, ENBIS, Amsterdam, 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
Brigitte Bartsch-Spörl, BSR Consulting GmbH, Germany
Isabelle Bichindaritz, University of Washington, USA
Leon Bobrowski, Bialystok Technical University, Poland
Marc Bouillé, France Télécom, France
Henning Christiansen, Roskilde University, Denmark
Shirley Coleman, University of Newcastle, UK
Juan M. Corchado, Universidad de Salamanca, Spain
Jeroen de Bruin, Medical University of Vienna, Austria
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
VIII Organization

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

Additional Reviewers

Francoise Fessant Orange Labs, France
Vincent Lemaire Orange Labs, France
Fabrice Clerot Orange Labs, France
Carine Hue Orange Labs, France
Dominique Gay Orange Labs, France
# Table of Contents

## Theoretical Aspects of Data Mining

- **Improvements over Adaptive Local Hyperplane to Achieve Better Classification**
  - Hongmin Cai
  - Page 1

- **Prognostic Models Based on Linear Separability**
  - Leon Bobrowski
  - Page 11

- **One Class Classification for Anomaly Detection: Support Vector Data Description Revisited**
  - Eric J. Pauwels and Onkar Ambekar
  - Page 25

- **How to Interpret Decision Trees?**
  - Petra Perner
  - Page 40

- **Comparing Classifiers and MetaClassifiers**
  - Elio Lozano and Edgar Acuña
  - Page 56

- **Fast Data Acquisition in Cost-Sensitive Learning**
  - Victor S. Sheng
  - Page 66

## Data Mining in Medicine

- **Application of a Unified Medical Data Miner (UMDM) for Prediction, Classification, Interpretation and Visualization on Medical Datasets: The Diabetes Dataset Case**
  - Nawaz Mohamudally and Dost Muhammad Khan
  - Page 78

- **Melanoma Diagnosis and Classification Web Center System: The Non-invasive Diagnosis Support Subsystem**
  - Wieslaw Paja and Mariusz Wrzesień
  - Page 96

- **Characterizing Cell Types through Differentially Expressed Gene Clusters Using a Model-Based Approach**
  - Juliane Perner and Elena Zotenko
  - Page 106

- **Experiments with Hybridization and Optimization of the Rules Knowledge Base for Classification of MMPI Profiles**
  - Jerzy Gomuła, Wiesław Paja, Krzysztof Pancerz, Teresa Mroczek, and Mariusz Wrzesień
  - Page 121
Multimedia Data Mining

Unsupervised Classification of Hyperspectral Images on Spherical Manifolds .................................................... 134
  *Dalton Lunga and Okan Ersoy*

Recognition of Porosity in Wood Microscopic Anatomical Images........ 147
  *Shen Pan and Mineichi Kudo*

Data Mining in Agriculture

Exploratory Hierarchical Clustering for Management Zone Delineation in Precision Agriculture ................................. 161
  *Georg Ruß and Rudolf Kruse*

High Classification Rates for Continuous Cow Activity Recognition Using Low-Cost GPS Positioning Sensors and Standard Machine Learning Techniques ............................................................. 174
  *Torben Godsk and Mikkel Baun Kjærgaard*

Mining Pixel Evolutions in Satellite Image Time Series for Agricultural Monitoring .......................................................... 189
  *Andreea Julea, Nicolas Méger, Christophe Rigotti, Emmanuel Trouvé, Philippe Bolon, and Vasile Lăzărescu*

Data Mining for Industrial Processes

Robust, Non-Redundant Feature Selection for Yield Analysis in Semiconductor Manufacturing ........................................... 204
  *Eric St. Pierre and Eugene Tuv*

Integrated Use of ICA and ANN to Recognize the Mixture Control Chart Patterns in a Process ........................................ 218
  *Yuehjen E. Shao, Yini Lin, and Ya-Chi Chan*

Optimized Fuzzy Decision Tree Data Mining for Engineering Applications ........................................................................... 228
  *Liam Evans and Niels Lohse*

Data Warehousing

Graph-Based Data Warehousing Using the Core-Facets Model ........... 240
  *Dung N. Lam, Alexander Y. Liu, and Cheryl E. Martin*
# Data Mining in Marketing

General Sales Forecast Models for Automobile Markets Based on Time Series Analysis and Data Mining Techniques ........................................ 255  
*Marco Hülsmann, Detlef Borscheid, Christoph M. Friedrich, and Dirk Reith*

# WebMining/Information Mining

Towards a Spatial Instance Learning Method for Deep Web Pages ...... 270  
*Ermelinda Oro and Massimo Ruffolo*

# Data Mining in Telecommunications

Applying User Signatures on Fraud Detection in Telecommunications Networks ................................................................. 286  
*João Lopes, Orlando Belo, and Carlos Vieira*

# Aspects of Data Mining

Methods in Case-Based Classification in Bioinformatics: Lessons Learned ................................................................. 300  
*Isabelle Bichindaritz*

Prediction of Batch-End Quality for an Industrial Polymerization Process ................................................................. 314  
*Geert Gins, Bert Pluymers, Ilse Y. Smets, Jairo Espinosa, and Jan F.M. Van Impe*

# Author Index

................................................................. 329