

Lecture Notes in Artificial Intelligence 6171

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Petra Perner (Ed.)

Advances in Data Mining

Applications and Theoretical Aspects

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



Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada

Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editor

Petra Perner

Institute of Computer Vision
and Applied Computer Sciences, IBaI
Kohlenstr. 2
04107 Leipzig, Germany
E-mail: pperner@ibai-institut.de

Library of Congress Control Number: 2010930175

CR Subject Classification (1998): I.2.6, I.2, H.2.8, J.3, H.3, I.4-5, J.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-642-14399-7 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-14399-1 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2010

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

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 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).

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
Andrea Ahlemeyer-Stubbe
Klaus-Dieter Althoff
Chid Apté
Eva Armengol
Bart Baesens
Isabelle Bichindaritz
Leon Bobrowski
Marc Boullé
Henning Christiansen
Shirley Coleman
Juan M. Corchado
Antonio Dourado
Peter Funk
Brent Gordon
Gary F. Holness
Eyke Hüllermeier
Piotr Jedrzejowicz
Janusz Kacprzyk
Mehmed Kantardzic
Ron Kenett
Mineichi Kudo
David Manzano Macho
Eduardo F. Morales
Stefania Montani
Jerry Oglesby
Eric Pauwels
Mykola Pechenizkiy

Ashwin Ram
Tim Rey
Rainer Schmidt
Yuval Shahar
David Taniar

Medical University of Vienna, Austria
ENBIS, The Netherlands
University of Hildesheim, Germany
IBM Yorktown Heights, USA
IIA CSIC, Spain
KU Leuven, Belgium
University of Washington, USA
Bialystok Technical University, Poland
France Télécom, France
Roskilde University, Denmark
University of Newcastle, UK
Universidad de Salamanca, Spain
University of Coimbra, Portugal
Mälardalen University, Sweden
NASA Goddard Space Flight Center, USA
Quantum Leap Innovations Inc., USA
University of Marburg, Germany
Gdynia Maritime University, Poland
Polish Academy of Sciences, Poland
University of Louisville, USA
KPA Ltd., Israel
Hokkaido University, Japan
Ericsson Research Spain, Spain
INAOE, Ciencias Computacionales, Mexico
Università del Piemonte Orientale, Italy
SAS Institute Inc., USA
CWI Utrecht, The Netherlands
Eindhoven University of Technology,
The Netherlands
Georgia Institute of Technology, USA
Dow Chemical Company, USA
University of Rostock, Germany
Ben Gurion University, Israel
Monash University, Australia

VIII Organization

Stijn Viaene
Rob A. Vingerhoeds
Yanbo J. Wang

Claus Weihns
Terry Windeatt

KU Leuven, Belgium
Ecole Nationale d'Ingénieurs de Tarbes, France
Information Management Center, China
Minsheng Banking Corporation Ltd., China
University of Dortmund, Germany
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	1
<i>Giorgio Giacinto</i>	

Bioinformatics Contributions to Data Mining	17
<i>Isabelle Bichindaritz</i>	

Theoretical Aspects of Data Mining

Bootstrap Feature Selection for Ensemble Classifiers	28
<i>Rakkrit Duangsoithong and Terry Windeatt</i>	

Evaluating the Quality of Clustering Algorithms Using Cluster Path Lengths	42
<i>Faraz Zaidi, Daniel Archambault, and Guy Melançon</i>	

Finding Irregularly Shaped Clusters Based on Entropy	57
<i>Angel Kuri-Morales and Edwin Aldana-Bobadilla</i>	

Fuzzy Conceptual Clustering	71
<i>Petra Perner and Anja Attig</i>	

Mining Concept Similarities for Heterogeneous Ontologies	86
<i>Konstantin Todorov, Peter Geibel, and Kai-Uwe Kühnberger</i>	

Re-mining Positive and Negative Association Mining Results	101
<i>Ayhan Demiriz, Gurdal Ertek, Tankut Atan, and Ufuk Kula</i>	

Multi-Agent Based Clustering: Towards Generic Multi-Agent Data Mining	115
<i>Santhana Chaimontree, Katie Atkinson, and Frans Coenen</i>	

Describing Data with the Support Vector Shell in Distributed Environments	128
<i>Peng Wang and Guojun Mao</i>	

Robust Clustering Using Discriminant Analysis	143
<i>Vasudha Bhatnagar and Sangeeta Ahuja</i>	

New Approach in Data Stream Association Rule Mining Based on Graph Structure	158
<i>Samad Gahderi Mojaveri, Esmaeil Mirzaeian, Zarrintaj Bornae, and Saeed Ayat</i>	

Multimedia Data Mining

Fast Training of Neural Networks for Image Compression	165
<i>Yevgeniy Bodyanskiy, Paul Grimm, Sergey Mashtalir, and Vladimir Vinarski</i>	

Processing Handwritten Words by Intelligent Use of OCR Results	174
<i>Benjamin Mund and Karl-Heinz Steinke</i>	

Saliency-Based Candidate Inspection Region Extraction in Tape Automated Bonding	186
<i>Martina Dümcke and Hiroki Takahashi</i>	

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

Entropic Quadtrees and Mining Mars Craters	210
<i>Rosanne Vetro and Dan A. Simovici</i>	

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

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

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

Data Mining in Marketing

Quantile Regression Model for Impact Toughness Estimation	263
<i>Satu Tamminen, Ilmari Juutilainen, and Juha Röning</i>	

Mining for Paths in Flow Graphs	277
<i>Adam Jocksch, José Nelson Amaral, and Marcel Mitran</i>	

Combining Unsupervised and Supervised Data Mining Techniques for Conducting Customer Portfolio Analysis	292
<i>Zhiyuan Yao, Annika H. Holmbom, Tomas Eklund, and Barbro Back</i>	
Managing Product Life Cycle with MultiAgent Data Mining System	308
<i>Serge Parshutin</i>	
Modeling Pricing Strategies Using Game Theory and Support Vector Machines	323
<i>Cristián Bravo, Nicolás Figueira, and Richard Weber</i>	

Data Mining in Industrial Processes

Determination of the Fault Quality Variables of a Multivariate Process Using Independent Component Analysis and Support Vector Machine	338
<i>Yuehjen E. Shao, Chi-Jie Lu, and Yu-Chiun Wang</i>	
Dynamic Pattern Extraction of Parameters in Laser Welding Process ...	350
<i>Gissel Velarde and Christian Binroth</i>	
Trajectory Clustering for Vibration Detection in Aircraft Engines	362
<i>Aurélien Hazan, Michel Verleysen, Marie Cottrell, and Jérôme Lacaille</i>	
Episode Rule-Based Prognosis Applied to Complex Vacuum Pumping Systems Using Vibratory Data.....	376
<i>Florent Martin, Nicolas Méger, Sylvie Galichet, and Nicolas Becourt</i>	
Predicting Disk Failures with HMM- and HSMM-Based Approaches	390
<i>Ying Zhao, Xiang Liu, Siqing Gan, and Weimin Zheng</i>	
Aircraft Engine Health Monitoring Using Self-Organizing Maps	405
<i>Etienne Côme, Marie Cottrell, Michel Verleysen, and Jérôme Lacaille</i>	

Data Mining in Medicine

Finding Temporal Patterns in Noisy Longitudinal Data: A Study in Diabetic Retinopathy	418
<i>Vassiliki Somarakis, Deborah Broadbent, Frans Coenen, and Simon Harding</i>	
Selection of High Risk Patients with Ranked Models Based on the CPL Criterion Functions	432
<i>Leon Bobrowski</i>	
Medical Datasets Analysis: A Constructive Induction Approach	442
<i>Wiesław Paja and Mariusz Wrzesień</i>	

Data Mining in Agriculture

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

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

WebMining

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

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

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

Interest-Determining Web Browser	518
<i>Khaled Bashir Shaban, Joannes Chan, and Raymond Szeto</i>	

Web-Site Boundary Detection	529
<i>Ayesh Alshukri, Frans Coenen, and Michele Zito</i>	

Data Mining in Finance

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

A Semi-supervised Approach for Reject Inference in Credit Scoring Using SVMs	558
<i>Sebastián Maldonado and Gonzalo Paredes</i>	

Aspects of Data Mining

Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool.....	572
<i>Paulo Cortez</i>	

The Orange Customer Analysis Platform	584
<i>Raphaël Féraud, Marc Boullé, Fabrice Clérot, Françoise Fessant, and Vincent Lemaire</i>	
Semi-supervised Learning for False Alarm Reduction	595
<i>Chien-Yi Chiu, Yuh-Jye Lee, Chien-Chung Chang, Wen-Yang Luo, and Hsiu-Chuan Huang</i>	
Learning from Humanoid Cartoon Designs	606
<i>Md. Tanvirul Islam, Kaiser Md. Nahiduzzaman, Why Yong Peng, and Golam Ashraf</i>	
Mining Relationship Associations from Knowledge about Failures Using Ontology and Inference	617
<i>Weisen Guo and Steven B. Kraines</i>	

Data Mining for Network Performance Monitoring

Event Prediction in Network Monitoring Systems: Performing Sequential Pattern Mining in Osmius Monitoring Tool	632
<i>Rafael García, Luis Llana, Constantino Malagón, and Jesús Pancorbo</i>	
Selection of Effective Network Parameters in Attacks for Intrusion Detection	643
<i>Gholam Reza Zargar and Peyman Kabiri</i>	

Author Index	653
---------------------------	-----