

Lecture Notes in Artificial Intelligence 5633

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

9th Industrial Conference, ICDM 2009
Leipzig, Germany, July 20-22, 2009
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: Applied for

CR Subject Classification (1998): I.2.6, I.2, H.2.8, K.4.4, J.3, I.4, J.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-642-03066-1 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-03066-6 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 2009

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper SPIN: 12718375 06/3180 5 4 3 2 1 0

Preface

This volume comprises the proceedings of the Industrial Conference on Data Mining (ICDM 2009) held in Leipzig (www.data-mining-forum.de).

For this edition the Program Committee received 130 submissions. After the peer-review process, we accepted 32 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.

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

In conjunction with ICDM two workshops were run focusing on special hot application-oriented topics in data mining. The workshop Data Mining in Marketing DMM 2009 was run for the second time. The papers are published in a separate workshop book “Advances in Data Mining on Marketing” by *ibai-publishing* (www.ibai-publishing.org). The Workshop on Case-Based Reasoning for Multimedia Data CBR-MD ran for the second year. The papers are published in a special issue of the *International Journal of Transactions on Case-Based Reasoning* (www.ibai-publishing.org/journal/cbr).

We are pleased to announce that we gave out the best paper award for ICDM fourth time. More details are mentioned at www.data-mining-forum.de. 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 a session on new challenging topics in data mining before the Best Paper Award Ceremony.

We also 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 ICDM will take place in Berlin in 2010.

Industrial Conference on Data Mining, ICDM 2009

Chair

Petra Perner

IBaI Leipzig, Germany

Committee

Klaus-Peter Adlassnig
Andrea Ahlemeyer-Stubbe

Klaus-Dieter Althoff
Chid Apte

Eva Armengol
Bart Baesens

Isabelle Bichindaritz
Leon Bobrowski

Marc Boullé
Henning Christiansen

Shirley Coleman
Juan M. Corchado

Da Deng
Antonio Dourado

Peter Funk
Brent Gordon

Gary F. Holness
Eyke Hüllermeier

Piotr Jedrzejowicz
Janusz Kacprzyk

Mehmed Kantardzic
Ron Kenett

Mineichi Kudo
Eduardo F. Morales

Stefania Montani
Jerry Oglesby

Eric Pauwels
Mykola Pechenizkiy

Ashwin Ram
Tim Rey
Rainer Schmidt
Yuval Shahar
David Tanian

Medical University of Vienna, Austria
ENBIS, Amsterdam

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 Otago, New Zealand
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
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

Stijn Viaene
Rob A. Vingerhoeds
Claus Weihs
Terry Windeatt

KU Leuven, Belgium
Ecole Nationale d'Ingénieurs de Tarbes, France
University of Dortmund, Germany
University of Surrey, UK

Table of Contents

Invited Talk

Distances in Classification.....	1
<i>Claus Weihs and Gero Szepannek</i>	

Data Mining in Medicine and Agriculture

Electronic Nose Ovarian Carcinoma Diagnosis Based on Machine Learning Algorithms	13
<i>José Chilo, György Horvath, Thomas Lindblad, and Roland Olsson</i>	
Data Mining of Agricultural Yield Data: A Comparison of Regression Models	24
<i>Georg Ruß</i>	
Study of Principal Components on Classification of Problematic Wine Fermentations	38
<i>Alejandra Urtubia U. and J. Ricardo Pérez-Correa</i>	
A Data Mining Method for Finding Hidden Relationship in Blood and Urine Examination Items for Health Check	44
<i>Kazuhiko Shinozawa, Norihiro Hagita, Michiko Furutani, and Rumiko Matsuoka</i>	
Application of Classification Association Rule Mining for Mammalian Mesenchymal Stem Cell Differentiation	51
<i>Weiqi Wang, Yanbo J. Wang, René Bañares-Alcántara, Zhanfeng Cui, and Frans Coenen</i>	
Computer-Aided Diagnosis in Brain Computed Tomography Screening	62
<i>Hugo Peixoto and Victor Alves</i>	
Knowledge Representation in Difficult Medical Diagnosis	73
<i>Ana Aguilera and Alberto Subero</i>	

Data Mining in Marketing, Finance and Telecommunication

Forecasting Product Life Cycle Phase Transition Points with Modular Neural Networks Based System	88
<i>Serge Parshutin, Ludmila Aleksejeva, and Arkady Borisov</i>	

Visualizing the Competitive Structure of Online Auctions	103
<i>Stephen France and Douglas Carroll</i>	
Credit Risk Handling in Telecommunication Sector	117
<i>Monika Szczerba and Andrzej Ciemski</i>	
Sales Intelligence Using Web Mining	131
<i>Viara Popova, Robert John, and David Stockton</i>	
A Sales Forecast Model for the German Automobile Market Based on Time Series Analysis and Data Mining Methods	146
<i>Bernhard Brühl, Marco Hülsmann, Detlef Borscheid, Christoph M. Friedrich, and Dirk Reith</i>	

Data Mining in Process Control, Industry and Society

Screening Paper Runnability in a Web-Offset Pressroom by Data Mining	161
<i>A. Alzghoul, A. Verikas, M. Hållander, M. Bacauskiene, and A. Gelzinis</i>	
Evaluation of Distraction in a Driver-Vehicle-Environment Framework: An Application of Different Data-Mining Techniques	176
<i>Fabio Tango and Marco Botta</i>	
SO_MAD: SensOr Mining for Anomaly Detection in Railway Data	191
<i>Julien Rabatel, Sandra Bringay, and Pascal Poncelet</i>	
Online Mass Flow Prediction in CFB Boilers	206
<i>Andriy Ivannikov, Mykola Pechenizkiy, Jorn Bakker, Timo Leino, Mikko Jegoroff, Tommi Kärkkäinen, and Sami Äyrämö</i>	
Integrating Data Mining and Agent Based Modeling and Simulation	220
<i>Omar Baqueiro, Yanbo J. Wang, Peter McBurney, and Frans Coenen</i>	
Combining Multidimensional Scaling and Computational Intelligence for Industrial Monitoring	232
<i>António Dourado, Sara Silva, Lara Aires, and João Araújo</i>	
A Case of Using Formal Concept Analysis in Combination with Emergent Self Organizing Maps for Detecting Domestic Violence	247
<i>Jonas Poelmans, Paul Elzinga, Stijn Viaene, and Guido Dedene</i>	

Data Mining on Multimedia Data

Ordinal Evaluation: A New Perspective on Country Images	261
<i>Marko Robnik-Šikonja, Kris Brijs, and Koen Vanhoof</i>	

Evaluation of Fusion for Similarity Searching in Online Handwritten Documents	276
<i>Sascha Schimke, Maik Schott, Claus Vielhauer, and Jana Dittmann</i>	
Self-training Strategies for Handwriting Word Recognition	291
<i>Volkmar Frinken and Horst Bunke</i>	
On a New Similarity Analysis in Frequency Domain for Mining Faces within a Complex Background	301
<i>D.A. Karras</i>	

Theoretical Aspects of Data Mining

Clustering with Domain Value Dissimilarity for Categorical Data	310
<i>Jeonghoon Lee, Yoon-Joon Lee, and Minho Park</i>	
The Normalized Compression Distance as a Distance Measure in Entity Identification	325
<i>Sebastian Klenk, Dennis Thom, and Gunther Heidemann</i>	
Attribute Constrained Rules for Partially Labeled Sequence Completion	338
<i>Chad A. Williams, Peter C. Nelson, and Abolfazl (Kouros) Mohammadian</i>	
Mining Determining Sets for Partially Defined Functions	353
<i>Dan A. Simovici, Dan Pletea, and Rosanne Vetro</i>	
On the Integration of Neural Classifiers through Similarity Analysis of Higher Order Features	361
<i>D.A. Karras and B.G. Mertzios</i>	
On Cellular Network Channels Data Mining and Decision Making through Ant Colony Optimization and Multi Agent Systems Strategies	372
<i>P.M. Papazoglou, D.A. Karras, and R.C. Papademetriou</i>	
Responsible Data Releases	388
<i>Sanguthevar Rajasekaran, Ofer Harel, Michael Zuba, Greg Matthews, and Robert Aseltine</i>	
Author Index	401