

# Lecture Notes in Artificial Intelligence 7377

## Subseries of Lecture Notes in Computer Science

### LNAI Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

### LNAI Founding Series Editor

Joerg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*

Petra Perner (Ed.)

# Advances in Data Mining

Applications and  
Theoretical Aspects

12th Industrial Conference, ICDM 2012  
Berlin, Germany, July 13-20, 2012  
Proceedings

## 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  
Kohlenstraße 2  
04107 Leipzig, Germany  
E-mail: pperner@ibai-institut.de

ISSN 0302-9743 e-ISSN 1611-3349  
ISBN 978-3-642-31487-2 e-ISBN 978-3-642-31488-9  
DOI 10.1007/978-3-642-31488-9  
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012940664

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

© Springer-Verlag Berlin Heidelberg 2012

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.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

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

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

The 12th event of the Industrial Conference on Data Mining ICDM was held in Berlin ([www.data-mining-forum.de](http://www.data-mining-forum.de)) running under the umbrella of the World Congress “The Frontiers in Intelligent Data and Signal Analysis, DSA 2012” ([www.worldcongressdsa.com](http://www.worldcongressdsa.com)).

For this edition the Program Committee received 97 submissions. After the peer-review process, we accepted 32 high-quality papers for oral presentation of which 22 are included in this proceedings book. The topics range from theoretical aspects of data mining to applications of data mining such as in multimedia data, marketing, finance and telecommunications, medicine and agriculture, and 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](http://www.ibai-publishing.org/journal/mldm)).

Fifteen papers were selected for poster presentations and six for industry paper presentations, which are published in the *ICDM Poster and Industry Proceedings* by *ibai-publishing* ([www.ibai-publishing.org](http://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 (CBR-MD), and the Workshop Data Mining in Agriculture (DMA). All workshop papers are published in the *workshop proceedings* by *ibai-publishing* ([www.ibai-publishing.org](http://www.ibai-publishing.org)).

A tutorial on Data Mining, a tutorial on Case-Based Reasoning, a tutorial on Intelligent Image Interpretation and Computer Vision in Medicine, Biotechnology, Chemistry and Food Industry and a tutorial on Standardization in Immunofluorescence were held before the conference.

We were pleased to give out the the best paper award for the sixth time this year ([www.data-mining-forum.de](http://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 discussions with the auditorium. The ceremony took place at the end of the conference. This prize is sponsored by *ibai solutions* ([www.ibai-solutions.de](http://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 on new challenging topics in data mining before the Best Paper Award Ceremony.

We would like to thank the members of the Institute of Applied Computer Sciences, Leipzig, Germany ([www.ibai-institut.de](http://www.ibai-institut.de)), who handed 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. We hope to see you in 2013 in New York at the next World Congress on “The Frontiers in Intelligent Data and Signal Analysis, DSA2013” ([www.worldcongressdsa.com](http://www.worldcongressdsa.com)) that combines the following three 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.

July 2012

Petra Pernert

# Organization

## Industrial Conference on Data Mining, ICDM 2012

### Chair

Petra Perner

IBaI Leipzig, Germany

### Committee

Klaus-Peter Adlassnig	Medical University of Vienna, Austria
Andrea Ahlemeyer-Stubbe	ENBIS, The Netherlands
Klaus-Dieter Althoff	University of Hildesheim, Germany
Eva Armengol	IIA CSIC, Spain
Brigitte Bartsch-Spörl	BSR Consulting GmbH, Germany
Orlando Belo	University of Minho, Portugal
Isabelle Bichindaritz	University of Washington, USA
Leon Bobrowski	Bialystok Technical University, Poland
Marc Boullé	France Télécom, France
Shirley Coleman	University of Newcastle, UK
Juan M. Corchado	Universidad de Salamanca, Spain
Jeremiah Da Deng	University of Otago, New Zealand
Jeroen de Bruin	Medical University of Vienna, Austria
Antonio Dourado	University of Coimbra, Portugal
Peter Funk	Mälardalen University, Sweden
Gary F. Holness	Quantum Leap Innovations Inc., USA
Piotr Jedrzejowicz	Gdynia Maritime University, Poland
Janusz Kacprzyk	Polish Academy of Sciences, Poland
Mehmed Kantardzic	University of Louisville, USA
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
Wieslaw Paja	University of Information Technology and Management in Rzeszow, Poland
Eric Pauwels	CWI Utrecht, The Netherlands
Mykola Pechenizkiy	Eindhoven University of Technology, The Netherlands

## VIII Organization

Georg Ruß

Otto-von-Guericke-Universität Magdeburg,  
Germany

Rainer Schmidt

University of Rostock, Germany

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

## Data Mining in Medicine and Biology

Application of Classification Algorithms on IDDM Rat Data .....	1
<i>Rainer Schmidt, Heike Weiss, and Georg Fuellen</i>	
Research Themes in the Case-Based Reasoning in Health Sciences Core Literature.....	9
<i>Isabelle Bichindaritz</i>	
Research on Application of Data Mining Methods to Diagnosing Gastric Cancer .....	24
<i>Arnis Kirshners, Serge Parshutin, and Marcis Leja</i>	
SOHAC: Efficient Storage of Tick Data That Supports Search and Analysis .....	38
<i>Gabor I. Nagy and Krisztian Buza</i>	

## Data Mining for Energy Industry

Electricity Consumption Time Series Profiling: A Data Mining Application in Energy Industry .....	52
<i>Hongyan Liu, Zhiyuan Yao, Tomas Eklund, and Barbro Back</i>	
Wind Turbines Fault Diagnosis Using Ensemble Classifiers .....	67
<i>Pedro Santos, Luisa F. Villa, Aníbal Reñones, Andrés Bustillo, and Jesús Maudes</i>	

## Data Mining in Traffic and Logistic

Bus Bunching Detection by Mining Sequences of Headway Deviations .....	77
<i>Luís Moreira-Matias, Carlos Ferreira, João Gama, João Mendes-Moreira, and Jorge Freire de Sousa</i>	

## Data Mining in Telecommunication

Detecting Abnormal Patterns in Call Graphs Based on the Aggregation of Relevant Vertex Measures .....	92
<i>Ronnie Alves, Pedro Ferreira, Joel Ribeiro, and Orlando Belo</i>	



## Data Mining in Engineering

Real-Time Mass Flow Estimation in Circulating Fluidized Bed .....	103
<i>Andriy Ivannikov, Mikko Jegeroff, and Tommi Kärkkäinen</i>	
Representation in Case-Based Reasoning Applied to Control Reconfiguration .....	113
<i>Ons Lejri and Moncef Tagina</i>	
The Influence of Input and Output Measurement Noise on Batch-End Quality Prediction with Partial Least Squares .....	121
<i>Jef Vanlaer, Pieter Van den Kerkhof, Geert Gins, and Jan F.M. Van Impe</i>	

## Theory in Data Mining

An Evolving Associative Classifier for Incomplete Database .....	136
<i>Kaoru Shimada</i>	
Improving Classifier Performance by Knowledge-Driven Data Preparation .....	151
<i>Laura Welcker, Stephan Koch, and Frank Dellmann</i>	
CWFM: Closed Contingency Weighted Frequent Itemsets Mining .....	166
<i>Eunkyoung Park, Younghee Kim, Ieejoon Kim, Jaeyeol Yoon, Jiyeon Lim, and Ungmo Kim</i>	
Prognostic Modeling with High Dimensional and Censored Data .....	178
<i>Leon Bobrowski and Tomasz Lukaszuk</i>	

## Theory in Data Mining: Clustering

<i>SHACUN</i> : Semi-supervised Hierarchical Active Clustering Based on Ranking Constraints .....	194
<i>Eya Ben Ahmed, Ahlem Nabli, and Faïez Gargouri</i>	
A Minimum Spanning Tree-Inspired Clustering-Based Outlier Detection Technique .....	209
<i>Xiaochun Wang, Xia Li Wang, and D. Mitch Wilkes</i>	

## Theory in Data Mining: Association Rule Mining and Decision Rule Mining

ML-DS: A Novel Deterministic Sampling Algorithm for Association Rules Mining .....	224
<i>Samir A. Mohamed Elsayed, Sanguthevar Rajasekaran, and Reda A. Ammar</i>	

Decision Rules Development Using Set of Generic Operations Approach .....	236
<i>Wiesław Paja and Mariusz Wrzesień</i>	

## Text Mining

Redundant Dictionary Spaces as a General Concept for the Analysis of Non-vectorial Data .....	243
<i>Sebastian Klenk, Jürgen Dippon, Andre Burkowski, and Gunther Heidemann</i>	
Human-Centered Text Mining: A New Software System .....	258
<i>Jonas Poelmans, Paul Elzinga, Alexei A. Neznanov, Guido Dedene, Stijn Viaene, and Sergei O. Kuznetsov</i>	
Text Mining Scientific Papers: A Survey on FCA-Based Information Retrieval Research .....	273
<i>Jonas Poelmans, Dmitry I. Ignatov, Stijn Viaene, Guido Dedene, and Sergei O. Kuznetsov</i>	
<b>Author Index</b> .....	289