

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

More information about this series at <http://www.springer.com/series/1244>

Petra Perner (Ed.)

Advances in Data Mining

Applications and Theoretical Aspects

16th Industrial Conference, ICDM 2016
New York, NY, USA, July 13–17, 2016
Proceedings

Editor
Petra Perner
Institute of Computer Vision
and applied Computer Sciences, IBaI
Leipzig, Saxony
Germany

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-41560-4 ISBN 978-3-319-41561-1 (eBook)
DOI 10.1007/978-3-319-41561-1

Library of Congress Control Number: 2016942891

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, 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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

The 16th 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 on “The Frontiers in Intelligent Data and Signal Analysis, DSA 2016” (www.worldcongressdsa.com).

After a peer-review process, we accepted 32 high-quality papers for oral presentation. The topics range from theoretical aspects of data mining to applications of data mining, such as in multimedia data, in marketing, in medicine, 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).

In all, ten 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).

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 best paper award for ICDM for the eighth time this year (three announcements 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 during 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.

We would like to thank all reviewers for their highly professional work and their effort in reviewing the papers.

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. We hope to see you in 2017 in New York at the next World Congress on “The Frontiers in Intelligent Data and Signal Analysis, DSA 2017” (www.worldcongressdsa.com), which combines under its roof the following three events: International Conferences 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).

Organizations

Chair

Petra Perner IBAI Leipzig, Germany

Program Committee

Ajith Abraham	Machine Intelligence Research Labs, USA
Brigitte Bartsch-Spörl	BSR Consulting GmbH, Germany
Orlando Belo	University of Minho, Portugal
Shirley Coleman	University of Newcastle, UK
Jeroen de Bruin	Medical University of Vienna, Austria
Antonio Dourado	University of Coimbra, Portugal
Geert Gins	KU Leuven, Belgien
Warwick Graco	ATO, Australia
Aleksandra Gruca	Silesian University of Technology, Poland
Hartmut Ilgner	Council for Scientific and Industrial Research, South Africa
Pedro Isaías	Universidade Aberta, Portugal
Piotr Jedrzejowicz	Gdynia Maritime University, Poland
Martti Juhola	University of Tampere, Finland
Janusz Kacprzyk	Polish Academy of Sciences, Poland
Mehmed Kantardzic	University of Louisville, USA
Mineichi Kudo	Hokkaido University, Japan
Eduardo F. Morales	INAOE, Ciencias Computacionales, Mexico
Armand Prieditris	Newstar Labs, USA
Rainer Schmidt	University of Rostock, Germany
Victor Sheng	University of Central Arkansas, USA
Kaoru Shimada	Fukuoka Dental College, Japan
Gero Szepannek	Santander Consumer Bank, Germany

Additional Reviewers

Alessandro Mammana	Max Planck Institute for Molecular Genetics
Juliane Perner	Cancer Research UK Cambridge Institute

Contents

Induction of Model Trees for Predicting BOD in River Water: A Data Mining Perspective.	1
<i>J. Alamelu Mangai and Bharat B. Gulyani</i>	
Efficient Mining of High Average-Utility Itemsets with Multiple Minimum Thresholds.	14
<i>Jerry Chun-Wei Lin, Ting Li, Philippe Fournier-Viger, Tzung-Pei Hong, and Ja-Hwung Su</i>	
Data Mining in Medicine: Relationship of Scoliotic Spine Curvature to the Movement Sequence of Lateral Bending Positions	29
<i>Athena Jalalian, Francis E.H. Tay, and Gabriel Liu</i>	
A Data Mining-Based Approach for Exploiting the Characteristics of University Lecturers	41
<i>Thuc-Doan Do, Thuy-Van T. Duong, and Ngoc-Phien Nguyen</i>	
Incremental Response Modeling Based on Segmentation Approach Using Uplift Decision Trees	54
<i>Sankara Prasad Kondareddy, Shruti Agrawal, and Shishir Shekhar</i>	
PHM: Mining Periodic High-Utility Itemsets	64
<i>Philippe Fournier-Viger, Jerry Chun-Wei Lin, Quang-Huy Duong, and Thu-Lan Dam</i>	
Comparative Analysis of Levenberg-Marquardt and Bayesian Regularization Backpropagation Algorithms in Photovoltaic Power Estimation Using Artificial Neural Network	80
<i>Kian Jazayeri, Moein Jazayeri, and Sener Uysal</i>	
Data Mining on Divers Alert Network DSL Database: Classification of Divers	96
<i>Tamer Ozyigit, Cuneyt Yavuz, Massimo Pieri, S. Murat Egi, Bahar Egi, Corentin Alepe, Danilo Cialoni, and Alessandro Marroni</i>	
Multiple Health Phases Based Remaining Useful Lifetime Prediction on Bearings	110
<i>Junjie Chen, Xiaofeng Wang, Wenjing Zhou, Lei Zhang, and Fei Liu</i>	
Innovations in News Media: Crisis Classification System.	125
<i>David Kaczynski, Lisa Gandy, and Gongzhu Hu</i>	

Development of Issue Sets from Social Big Data: A Case Study of Green Energy and Low-Carbon	139
<i>Chun-Che Huang, Yu-Jie Fang, Shian-Hua Lin, Wen-Yau Liang, and Shu-Rong Wu</i>	
Early Prediction of Extreme Rainfall Events: A Deep Learning Approach . . .	154
<i>Sulagna Gope, Sudeshna Sarkar, Pabitra Mitra, and Subimal Ghosh</i>	
Identifying and Characterizing Truck Stops from GPS Data	168
<i>Russel Aziz, Manav Kedia, Soham Dan, Sayantan Basu, Sudeshna Sarkar, Sudeshna Mitra, and Pabitra Mitra</i>	
A Fuzzy Rule-Based Learning Algorithm for Customer Churn Prediction. . . .	183
<i>Bingquan Huang, Ying Huang, Chongcheng Chen, and M.-T. Kechadi</i>	
DDoS Attacks Detection in Cloud Computing Using Data Mining Techniques	197
<i>Konstantin Borisenko, Andrey Smirnov, Evgenia Novikova, and Andrey Shorov</i>	
Coordinate Refinement on All Atoms of the Protein Backbone with Support Vector Regression.	212
<i>Ding-Yao Huang, Chiou-Yi Hor, and Chang-Biau Yang</i>	
The Computational Wine Wheel 2.0 and the TriMax Triclustering in Wineinformatics	223
<i>Bernard Chen, Christopher Rhodes, Alexander Yu, and Valentin Velchev</i>	
Understanding the Wine Judges and Evaluating the Consistency Through White-Box Classification Algorithms	239
<i>Bernard Chen, Hai Le, Christopher Rhodes, and Dongsheng Che</i>	
A Novel Sparsity Based Classification Framework to Exploit Clusters in Data	253
<i>Sudarshan Babu</i>	
Mining Event Sequences from Social Media for Election Prediction	266
<i>Kuan-Chieh Tung, En Tzu Wang, and Arbee L.P. Chen</i>	
Multi-exponential Lifetime Extraction in Time-Logarithmic Scale	282
<i>Andrew V. Knyazev, Qun Gao, and Koon Hoo Teo</i>	
Fuel Pipeline Thermal Conductivity in Automatic Wet Stock Reconciliation Systems	297
<i>Pawel Foszner, Aleksandra Gruca, and Jakub Bularz</i>	

A Novel Multivariate Mapping Method for Analyzing High-Dimensional Numerical Datasets	311
<i>Edwin Aldana-Bobadilla and Alejandro Molina-Villegas</i>	
Association Rule Hiding for Privacy Preserving Data Mining	320
<i>Shyma Mogtaba and Eiman Kambal</i>	
Extending Process Monitoring to Simultaneous False Alarm Rejection and Fault Identification (FARFI)	334
<i>Geert Gins, Sam Wuyts, Sander Van den Zegel, and Jan Van Impe</i>	
Parallel and Distributed Data Mining in Cloud	349
<i>Ivan Kholod, Mikhail Kuprianov, and Ilya Petukhov</i>	
The WoMan Formalism for Expressing Process Models	363
<i>Stefano Ferilli</i>	
Duration-Aware Alignment of Process Traces	379
<i>Sen Yang, Moliang Zhou, Rachel Webman, JaeWon Yang, Aleksandra Sarcevic, Ivan Marsic, and Randall S. Burd</i>	
A Matching Approach Based on Term Clusters for eRecruitment	394
<i>Gülşen Bal, Aşkın Karakaş, Tunga Güngör, Fatmagül Süzen, and Kemal Can Kara</i>	
Splitting Algorithm for Detecting Structural Changes in Predictive Relationships	405
<i>Olga Gorskikh</i>	
Early Predictive System for Diabetes Mellitus Disease	420
<i>Karim M. Orabi, Yasser M. Kamal, and Thanaa M. Rabah</i>	
Data Quality Visualization for Preprocessing	428
<i>Markus Vattulainen</i>	
List Price Optimization Using Customized Decision Trees	438
<i>R. Kiran, Shashank Shekhar, John Kiran, Raghava Rau, Sam Pritchett, Anit Bhandari, and Parag Chitalia</i>	
Author Index	445