

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

17th Industrial Conference, ICDM 2017
New York, NY, USA, July 12–13, 2017
Proceedings

Editor
Petra Perner
Institute of Computer Vision and Applied
Computer Sciences
Leipzig, Sachsen
Germany

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-62700-7 ISBN 978-3-319-62701-4 (eBook)
DOI 10.1007/978-3-319-62701-4

Library of Congress Control Number: 2017945723

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing AG 2017

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. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, 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 2017” (www.worldcongressdsa.com).

After the peer-review process, we accepted 26 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).

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 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 2018 in New York at the next World Congress on “The Frontiers in Intelligent Data and Signal Analysis, DSA 2018” (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.

July 2017

Petra Perner

Organization

Chair

Petra Perner

IBaI Leipzig, Germany

Program Committee

Ajith Abraham

Brigitte Bartsch-Spörl

Orlando Belo

Bernard Chen

Jeroen de Bruin

Antonio Dourado

Stefano Ferilli

Geert Gins

Warwick Graco

Aleksandra Gruca

Hartmut Ilgner

Pedro Isaías

Piotr Jedrzejowicz

Martti Juhola

Janusz Kacprzyk

Mehmed Kantardzic

Eduardo F. Morales

Armand Prieditris

Rainer Schmidt

Victor Sheng

Kaoru Shimada

Gero Szepannek

Machine Intelligence Research Labs, USA

BSR Consulting GmbH, Germany

University of Minho, Portugal

University of Central Arkansas, USA

Medical University of Vienna, Austria

University of Coimbra, Portugal

University of Bari, Italy

KU Leuven, Belgium

ATO, Australia

Silesian University of Technology, Poland

Council for Scientific and Industrial Research,
South Africa

Universidade Aberta, Portugal

Gdynia Maritime University, Poland

University of Tampere, Finland

Polish Academy of Sciences, Poland

University of Louisville, USA

INAOE, Ciencias Computacionales, Mexico

Newstar Labs, USA

University of Rostock, Germany

University of Central Arkansas, USA

Section of Medical Statistics, Fukuoka Dental College,
Japan

Santander Consumer Bank, Germany

Additional Reviewer

Juliane Perner

Cancer Research UK Cambridge Institute

Contents

Incorporating Positional Information into Deep Belief Networks for Sentiment Classification	1
<i>Yong Jin, Harry Zhang, and Donglei Du</i>	
Tracking Multiple Social Media for Stock Market Event Prediction.	16
<i>Fang Jin, Wei Wang, Prithwish Chakraborty, Nathan Self, Feng Chen, and Naren Ramakrishnan</i>	
Ensemble Sales Forecasting Study in Semiconductor Industry.	31
<i>Qiuping Xu and Vikas Sharma</i>	
Towards an efficient method of modeling “Next Best Action” for Digital Buyer’s journey in B2B	45
<i>Anit Bhandari, Kiran Rama, Nandini Seth, Nishant Niranjana, Parag Chitalia, and Stig Berg</i>	
Association Rule-based Classifier Using Artificial Missing Values	57
<i>Kaoru Shimada, Takaaki Arahira, and Takashi Hanioka</i>	
Mining Location-based Service Data for Feature Construction in Retail Store Recommendation	68
<i>Tsung-Yi Chen, Lyu-Cian Chen, and Yuh-Min Chen</i>	
Constraint-based Clustering Algorithm for Multi-Density Data and Arbitrary Shapes	78
<i>Walid Atwa and Kan Li</i>	
Towards a Large Scale Practical Churn Model for Prepaid Mobile Markets. . .	93
<i>Amit Kumar Meher, Jobin Wilson, and R. Prashanth</i>	
Smart Stores: A scalable foot traffic collection and prediction system	107
<i>Soheila Abrishami, Piyush Kumar, and Wickus Nienaber</i>	
Multivariate Time Series Representation and Similarity Search Using PCA . . .	122
<i>Aminata Kane and Nematollaah Shiri</i>	
Message Passing on Factor Graph: A Novel Approach for Orphan Drug Physician Targeting.	137
<i>Yunlong Wang and Yong Cai</i>	
Fast GPU-based Influence Maximization within Finite Deadlines via Node-level Parallelism	151
<i>Koushik Pal, Zissis Poulos, Edward Kim, and Andreas Veneris</i>	

Visual Scenes Mining for Agent Awareness Module 166
Gang Ma, Zhentao Tang, Xi Yang, Zhongzhi Shi, and Kun Yang

Predicting Hospital Re-admissions from Nursing Care Data of Hospitalized Patients 181
Muhammad K. Lodhi, Rashid Ansari, Yingwei Yao, Gail M. Keenan, Diana Wilkie, and Ashfaq A. Khokhar

Activity Prediction in Process Management using the WoMan Framework . . . 194
Stefano Ferilli, Domenico Redavid, and Sergio Angelastro

Hierarchical Text Classification of Autopsy Reports to Determine MoD and CoD through Term-Based and Concepts-Based Features 209
Ghulam Mujtaba, Liyana Shuib, Ram Gopal Raj, Mohammed Ali Al-Garadi, Retnagowri Rajandram, and Khairunisa Shaikh

Collaborative Filtering Fusing Label Features Based on SDAE 223
Huan Huo, Xiufeng Liu, Deyuan Zheng, Zonghan Wu, Shengwei Yu, and Liang Liu

Interestingness Classification of Association Rules for Master Data 237
Wei Han, Julio Borges, Peter Neumayer, Yong Ding, Till Riedel, and Michael Beigl

MapReduce and Spark-based Analytic Framework Using Social Media Data for Earlier Flu Outbreak Detection. 246
Ali Al Essa and Miad Faezipour

An Integrated Approach using Data Mining and System Dynamics to Policy Design: Effects of Electric Vehicle Adoption on CO₂ Emissions in Singapore. 258
Bohao Zhang and Francis E.H. Tay

Using the Results of Capstone Analysis to Predict a Weather Outcome 269
Anthony G. Nolan and Warwick J. Graco

Classification of Network Traffic Using Fuzzy Clustering for Network Security 278
Terrence P. Fries

Machine Learning and Pattern Recognition Techniques for Information Extraction to Improve Production Control and Design Decisions. 286
Carlos A. Escobar and Ruben Morales-Menendez

Real-time Prediction of Styrene Production Volume based on Machine Learning Algorithms 301
Yikai Wu, Fang Hou, and Xiaopei Cheng

Contents	XI
A graph-based ranking model for automatic keyphrases extraction from Arabic documents	313
<i>Mohamed Salim El Bazzi, Driss Mammass, Taher Zaki, and Abdelatif Ennaji</i>	
Mining Frequent Subgraph Pattern over a Collection of Attributed-Graphs and Construction of a Relation Hierarchy for Result Reporting	323
<i>Petra Perner</i>	
Author Index	345