

Lecture Notes in Artificial Intelligence

10933

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

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

Petra Perner (Ed.)

Advances in Data Mining

Applications and Theoretical Aspects

18th Industrial Conference, ICDM 2018

New York, NY, USA, July 11–12, 2018

Proceedings

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

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-95785-2 ISBN 978-3-319-95786-9 (eBook)
<https://doi.org/10.1007/978-3-319-95786-9>

Library of Congress Control Number: 2018947574

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing AG, part of Springer Nature 2018

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 the registered company Springer International Publishing AG
part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The 18th event of the Industrial Conference on Data Mining ICDM was held in New York again (www.data-mining-forum.de) under the umbrella of the World Congress on Frontiers in Intelligent Data and Signal Analysis, DSA 2018 (www.worldcongressdsa.com).

After the peer-review process, we accepted 25 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 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).

In all, 20 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).

The tutorial days rounded up the high quality of the conference. Researchers and practitioners got an excellent insight in the research and technology of the respective fields, the new trends, and the open research problems that we would like to study further.

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 and in between the conferences of DSA 2018.

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 2019 in New York at the next World Congress on Frontiers in Intelligent Data and Signal Analysis, DSA 2019 (www.worldcongressdsa.com), which combines under its roof the following three events: International Conferences Machine Learning and Data Mining, MLDM (www.mldm.de), the Industrial Conference on Data Mining, ICDM (www.data-mining-forum.de), and the International Conference on Mass Data Analysis of Signals and Images in Medicine, Biotechnology, Chemistry, Biometry, Security, Agriculture, Drug Discovery and Food Industry, MDA (www.mda-signals.de), as well as the workshops, and tutorials.

Organization

Chair

Petra Perner

IBal Leipzig, Germany

Program Committee

Ajith Abraham

Machine Intelligence Research Labs (MIR Labs), USA

Brigitte Bartsch-Spörl

BSR Consulting GmbH, Germany

Orlando Belo

University of Minho, Portugal

Bernard Chen

University of Central Arkansas, USA

Antonio Dourado

University of Coimbra, Portugal

Jeroen de Bruin

Medical University of Vienna, Austria

Stefano Ferilli

University of Bari, Italy

Geert Gins

KU Leuven, Belgium

Warwick Graco

ATO, Australia

Aleksandra Gruca

Silesian University of Technology, Poland

Hartmut Ilgner

Council for Scientific and Industrial Research,
South Africa

Pedro Isaias

Universidade Aberta (Portuguese Open University),
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

Eduardo F. Morales

INAOE, Ciencias Computacionales, Mexico

Samuel Noriega

Universitat de Barcelona Spain

Juliane Perner

Cancer Research, Cambridge Institutes, UK

Armand Prieditris

Newstar Labs, USA

Rainer Schmidt

University of Rostock, Germany

Victor Sheng

University of Central Arkansas, USA

Kaoru Shimada

Section of Medical Statistics, Fukuoka Dental College,
Japan

Gero Szepannek

Stralsund University, Germany

Markus Vattulainen

Tampere University, Finland

Additional Reviewers

Dimitrios Karras
Calin Ciufudean
Valentin Brimkov
Michelangelo Ceci
Reneta Barneva
Christoph F. Eick
Thang Pham
Giorgio Giacinto
Kamil Dimililer

Contents

An Adaptive Oversampling Technique for Imbalanced Datasets	1
<i>Shaukat Ali Shahee and Usha Ananthakumar</i>	
From Measurements to Knowledge - Online Quality Monitoring and Smart Manufacturing	17
<i>Satu Tamminen, Henna Tiensuu, Eija Ferreira, Heli Helakoski, Vesa Kyllönen, Juha Jokisaari, and Esa Puukko</i>	
Mining Sequential Correlation with a New Measure	29
<i>Mohammad Fahim Arefin, Maliha Tashfia Islam, and Chowdhury Farhan Ahmed</i>	
A New Approach for Mining Representative Patterns	44
<i>Abeda Sultana, Hosneara Ahmed, and Chowdhury Farhan Ahmed</i>	
An Effective Ensemble Method for Multi-class Classification and Regression for Imbalanced Data	59
<i>Tahira Alam, Chowdhury Farhan Ahmed, Sabit Anwar Zahin, Muhammad Asif Hossain Khan, and Maliha Tashfia Islam</i>	
Automating the Extraction of Essential Genes from Literature.	75
<i>Ruben Rodrigues, Hugo Costa, and Miguel Rocha</i>	
Rise, Fall, and Implications of the New York City Medallion Market	88
<i>Sherraina Song</i>	
An Intelligent and Hybrid Weighted Fuzzy Time Series Model Based on Empirical Mode Decomposition for Financial Markets Forecasting	104
<i>Ruixin Yang, Junyi He, Mingyang Xu, Haoqi Ni, Paul Jones, and Nagiza Samatova</i>	
Evolutionary DBN for the Customers' Sentiment Classification with Incremental Rules	119
<i>Ping Yang, Dan Wang, Xiao-Lin Du, and Meng Wang</i>	
Clustering Professional Baseball Players with SOM and Deciding Team Reinforcement Strategy with AHP.	135
<i>Kazuhiro Kohara and Shota Enomoto</i>	
Data Mining with Digital Fingerprinting - Challenges, Chances, and Novel Application Domains	148
<i>Matthias Vodel and Marc Ritter</i>	

Categorization of Patient Diseases for Chinese Electronic Health Record Analysis: A Case Study	162
<i>Junmei Zhong, Xiu Yi, De Xuan, and Ying Xie</i>	
Dynamic Classifier and Sensor Using Small Memory Buffers	173
<i>R. Gelbard and A. Khalemsky</i>	
Speeding Up Continuous kNN Join by Binary Sketches.	183
<i>Filip Nalepa, Michal Batko, and Pavel Zezula</i>	
Mining Cross-Level Closed Sequential Patterns.	199
<i>Rutba Aman and Chowdhury Farhan Ahmed</i>	
An Efficient Approach for Mining Weighted Sequential Patterns in Dynamic Databases	215
<i>Sabrina Zaman Ishita, Faria Noor, and Chowdhury Farhan Ahmed</i>	
A Decision Rule Based Approach to Generational Feature Selection	230
<i>Wiesław Paja</i>	
A Partial Demand Fulfilling Capacity Constrained Clustering Algorithm to Static Bike Rebalancing Problem.	240
<i>Yi Tang and Bi-Ru Dai</i>	
Detection of IP Gangs: Strategically Organized Bots	254
<i>Tianyue Zhao and Xiaofeng Qiu</i>	
Medical AI System to Assist Rehabilitation Therapy	266
<i>Takashi Isobe and Yoshihiro Okada</i>	
A Novel Parallel Algorithm for Frequent Itemsets Mining in Large Transactional Databases	272
<i>Huan Phan and Bac Le</i>	
A Geo-Tagging Framework for Address Extraction from Web Pages.	288
<i>Julia Efremova, Ian Endres, Isaac Vidas, and Ofer Melnik</i>	
Data Mining for Municipal Financial Distress Prediction	296
<i>David Alaminos, Sergio M. Fernández, Francisca García, and Manuel A. Fernández</i>	
Prefix and Suffix Sequential Pattern Mining	309
<i>Rina Singh, Jeffrey A. Graves, Douglas A. Talbert, and William Eberle</i>	
Author Index	325