

Petra Perner (Ed.)

Advances in Data Mining

Applications and Theoretical Aspects

19th Industrial Conference, ICDM 2019
New York, USA, July 17 – July 21, 2019
Proceedings

Volume Editor

Petra Perner
Institute of Computer Vision and Applied Computer Sciences, IBaI
PF 30 11 14
04251 Leipzig
E-mail: pperner@ibai-institut.de

ISSN 1864-9734
ISBN 978-3-942952-60-6



The German National Library listed this publication in the German National Bibliography.
Detailed bibliographical data can be downloaded from <http://dnb.ddb.de>.

ibai-publishing
Prof. Dr. Petra Perner
PF 30 11 38
04251 Leipzig, Germany
E-mail: info@ibai-publishing.org
<http://www.ibai-publishing.org>

Copyright © 2019 ibai-publishing
ISSN 1864-9734
ISBN 978-3-942952-60-6

All rights reserved.
Printed in Germany, 2019

Preface

The nineteenth 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 2019” (www.worldcongressdsa.com).

After the peer-review process, we accepted 39 high-quality papers for oral presentation, which are published in the ICDM Proceeding by ibai-publishing (www.ibai-publishing.org). 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, twenty one papers were selected for poster presentations, which are published in the ICDM Poster Proceeding by ibai-publishing (www.ibai-publishing.org).

A tutorial on Data Mining and a tutorial on Case-Based Reasoning were held after the conference.

The conference was running in an inspiring atmosphere. The presenters of the oral presentations gave excellent talks and the audience acknowledged each talk with excellent questions. The poster presenters presented well prepared posters and gave in front of their posters a summary of their work in five minutes for the audiences. Afterwards the audience could step to the poster that they were interested in. The audience had a lot of questions and they gave valuable comments and new directions for the work in progress. In all was the poster session a very successful session.

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 ibai-publishing publishing house (www.ibai-publishing.org) that handled the papers and published the proceedings.

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 2020 in New York at the next World Congress on “The Frontiers in Intelligent Data and Signal Analysis, DSA 2020” (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 Artificial Intelligence and Pattern Recognition with Applications in Medicine, Biotechnology, Chemistry and Food Industry, MDA-AI&PR (www.mda-signals.de).

July 2019

Petra Perner

19th Industrial Conference on Data Mining ICDM 2019

www.data-mining-forum.de
July 17 – 21, 2019, New York, USA

Chair

Prof. Dr. Petra Perner

Institute of Computer Vision and applied Computer Sciences, IBAI

Program Committee

Ajith Abraham	Machine Intelligence Research Labs (MIR Labs), USA
Mohamed, Bourguessa	Universite du Quebec a Montreal - UQAM, Canada
Bernard Chen	University of Central Arkansas, USA
Jeroen de Bruin	University of Applied Sciences JOANNEUM, Austria
Antonio Dourado	University of Coimbra, Portugal
Stefano Ferilli	University of Bari, Italy
Geert Gins	Glaxo Smith Kline, Belgium
Warwick Graco	Australian Tax Office ATO, Australia
Aleksandra Gruca	Silesian University of Technology, Poland
Pedro Isaias	The University of Queensland, Australia
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
Lui Xiaobing	Google Inc., USA
Eduardo F. Morales	National Institute of Astrophysics, Optics, and Electronics, Mexico
Samuel Noriega	Universitat de Barcelona, Spain
Wieslaw Paja	University of Rzeszow, Poland
Juliane Perner	Novartis Institutes for BioMedical Research (NIBR), Switzerland
Rainer Schmidt	University of Rostock, Germany
Moti Schneider	PCCW Global, Greece
Victor Sheng	University of Central Arkansas, USA
Kaoru Shimada	Fukuoka Dental College, Japan
Gero Szepannek	University of Applied Sciences Stralsund, Germany
Joao Miguel Costa Sousa	Technical University of Lisbon, Portugal
Markus Vattulainen	Tampere University, Finland
Zhu Bing	Sichuan University, China

Table of Content

Ensemble of Customized DenseNet with SVM for Refining Iris Contact Lens Features <i>Meenakshi Choudhary, Venkanna U. and Vivek Tiwari</i>	1
An efficient approach for mining weighted frequent patterns with dynamic weights <i>Umama Dewan, Chowdhury Farhan Ahmed, Carson K. Leung, Redwan Ahmed Rizvee, Deyu Deng and Joglas Souza</i>	13
How to Achieve High Classification Accuracy with Just a Few Labels: A Semi-supervised Approach Using Sampled Packets <i>Shahbaz Rezaei and Xin Liu</i>	28
FAT-DeepFFM: Field Attentive Deep Field-aware Factorization Machine <i>Junlin Zhang, Tongwen Huang and Zhiqi Zhang</i>	43
A Proportional Process Mining System <i>Dinh-Lam Pham, Hyun Ahn, Minjae Park, Kyoung-Sook Kim and Kwanghoon Pio Kim</i>	58
Improving document classification performance by combining resources and integrating word embedding techniques <i>Kazem Qazanfari and Abdou Youssef</i>	75
A Hybrid Model on Learning Cross Features for Transaction Fraud Detection <i>Han Wu and Guanjun Liu</i>	88
Efficiency assessment of event-based predictive maintenance in Industry 4.0 <i>Athanasios Naskos and Anastasios Gounaris</i>	103
Sliding window based weighted periodic pattern mining over time series data <i>Redwan Ahmed Rizvee, Md Shahadat Hossain Shahin, Chowdhury Farhan Ahmed, Carson K. Leung, Deyu Deng and Jiaying Jason Mai</i>	118
A Symmetry of the Data Pattern Structure <i>Alexey Myachin</i>	133
The Use of Frequent Subgraph Mining to Develop a Recommender System for Playing Real-Time Strategy Games <i>Isam Alobaidi, Jennifer Leopold and Ali Allami</i>	146

Anomaly Detection in Univariate Time Series: An Empirical Comparison of Machine Learning Algorithms <i>Sina Däubener, Sebastian Schmitt, Hao Wang, Peter Krause and Thomas Bäck</i>	161
Predictive Analysis of Real-Time Strategy Games Using Discriminative Subgraph Mining <i>Jennifer Leopold, Isam Alobaidi and Nathan Elou</i>	176
Interpretable detection of unstable smart TV usage from power state logs <i>Tamir Mazaev, Olivier Janssens, Dirk Van Gheel, Lieve Lanoye and Sofie Van Hoecke</i>	191
WeFreS: weighted frequent subgraph mining in a single large graph <i>Nahian Ashraf, Riddho Ridwanul Haque, Md. Ashraf Islam, Chowdhury Farhan Ahmed, Carson K. Leung, Jiaying Jason Mai and Bryan H. Wodi</i>	201
Predicting e-commerce customer conversion from minimal temporal patterns on symbolized clickstream trajectories <i>Jacopo Tagliabue, Lucas Lucasa, Ciro Greco, Mattia Pavoni and Andrea Polonioli</i>	216
Functions and Architecture for Automatic Predictive Targeting and Modelling Real-time On-line Behavior <i>Petra Pernert</i>	221
Cluster-Based Relative Outlier Under-Sampling Technique <i>Atif Hassan, Rohini Banerjee, Pabitra Mitra and Pralay Mitra</i>	229
Weather Impacts on Wine, A BiMax examination of Napa Cabernet in 2011 and 2012 Vintage <i>Bernard Chen, David Jones, Mustafa Tunc, Kristen Chipolla and Jorge Beltran</i>	242
Context Aware Image Annotation in Multiple-Instance Active Learning <i>Yingcheng Sun, Kenneth Loparo</i>	251
Risk and Error Matrix Charts <i>Hari Koesmarno</i>	263
Rapidly Determining the Starting Sample Size in Progressive Sampling: Mean Convergence is Sufficient <i>Amr ElRafey, Janusz Wojtusiak</i>	278

Series of Optimized Predictive Models for Forecast of Global Wastewater Treatment Plant Performance <i>Bharat B. Gulyani, Arshia Fathima</i>	290
Compositional Analysis for Metallic Systems: Sampling Methods <i>Harsha Gwalani, Martin O Neill II, Armin R Mikler and Talukder Alam, Rajarshi Banerjee</i>	303
Scalable Cloud-based ETL for Self-serving Analytics <i>Eftim Zdravevski, Cas Apanowicz, Krzysztof Stencel and Dominik Slezak</i>	317
An Exploration into the Contexts of Errors and Value Patterns in Data Classification <i>William Wu</i>	332
Authors Index	346