

Lecture Notes in Artificial Intelligence

12986

Subseries of Lecture Notes in Computer Science

Series Editors

Randy Goebel

University of Alberta, Edmonton, Canada

Yuzuru Tanaka

Hokkaido University, Sapporo, Japan

Wolfgang Wahlster

DFKI and Saarland University, Saarbrücken, Germany

Founding Editor

Jörg Siekmann

DFKI and Saarland University, Saarbrücken, Germany

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

Carlos Soares · Luis Torgo (Eds.)

Discovery Science

24th International Conference, DS 2021
Halifax, NS, Canada, October 11–13, 2021
Proceedings

Editors

Carlos Soares 
Universidade do Porto and Fraunhofer
Portugal AICOS
Porto, Portugal

Luis Torgo 
Dalhousie University
Halifax, NS, Canada

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-030-88941-8 ISBN 978-3-030-88942-5 (eBook)
<https://doi.org/10.1007/978-3-030-88942-5>

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer Nature Switzerland AG 2021

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, expressed 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.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains the papers selected for presentation at the 24th International Conference on Discovery Science (DS 2021), which was organized to be held in Halifax, Canada, during October 11–13, 2021. Due to the restrictions associated with the COVID-19 pandemic, the conference was moved online and held virtually over the same time period.

DS is a conference series that started in 1986. Held every year, DS continues its tradition as the unique venue for the latest advances in the development and analysis of methods for discovering scientific knowledge, coming from machine learning, data mining, and intelligent data analysis, along with their application in various scientific domains. In particular, major areas selected for DS 2021 included the following: applications (including a relevant number of papers addressing the COVID-19 pandemic), classification, data streams, feature selection, graph and network mining, neural networks and deep learning, preferences, recommender systems, representation learning, responsible artificial intelligence, and spatial, temporal and spatiotemporal data.

DS 2021 received 76 international submissions that were carefully reviewed by three or more Program Committee (PC) members or external reviewers, with a few exceptions. After a rigorous reviewing process, 15 regular papers and 21 short papers were accepted for presentation at the conference and publication in the DS 2021 volume.

We would like to sincerely thank all people who helped this volume come into being and made DS 2021 a successful and exciting event. In particular, we would like to express our appreciation for the work of the DS 2021 PC members and external reviewers who helped assure the high standard of accepted papers. We would like to thank all authors of submissions to DS 2021, without whose work it would not have been possible to have such high-quality contributions in the conference.

We are grateful to the Steering Committee chair, Sašo Džeroski, for his extraordinary support in critical decisions concerning the event plan, particularly important in these challenging times. We are also very grateful to the Program Committee chairs of DS 2020, Annalisa Appice and Grigorios Tsoumakas, for all the information provided, which made the whole organization much easier. We wish to express our thanks to the local organization chairs, David Langstroth, Nuno Moniz, Paula Branco, Vitor Cerqueira, and Yassine Baghoussi, for their support and incredible work. We would like to express our deepest gratitude to all those who served as organizers, session chairs, and hosts, who made great efforts to meet the online challenge to make the virtual conference a real success. Finally, our thanks are due to Alfred Hofmann and Anna Kramer of Springer for their continuous support and work on the proceedings. We are grateful to Springer for a special issue on Discovery Science to be published in the Machine Learning journal. All authors were given the possibility to extend and rework versions of their papers presented at DS 2021 for a chance to be published in this prestigious journal. For DS 2021, Springer also supported a Best Student Paper Award, which was given to Bart Bussmann and his co-authors, Jannes Nys and Steven Latré, for their paper “Neural Additive Vector Autoregression Models for Causal Discovery in Time Series.”

This paper presents high quality work on a very relevant topic and is complemented with the materials to reproduce it. We would like to congratulate the authors for this achievement.

September 2021

Carlos Soares
Luis Torgo

Organization

Program Committee Chairs

Carlos Soares	Universidade do Porto and Fraunhofer Portugal AICOS, Portugal
Luis Torgo	Dalhousie University, Canada

Program Committee

Martin Atzmueller	Tilburg University, The Netherlands
Colin Bellinger	National Research Council of Canada, Canada
Paula Branco	Ottawa University, Canada
Alberto Cano	Virginia Commonwealth University, USA
Daniel Castro Silva	University of Porto, Portugal
Michelangelo Ceci	University of Bari Aldo Moro, Italy
Victor Cerqueira	Dalhousie University, Canada
Bruno Cremilleux	University of Caen Normandy, France
Nicola Di Mauro	University of Bari Aldo Moro, Italy
Ivica Dimitrovski	Ss. Cyril and Methodius University in Skopje, North Macedonia
Wouter Duivesteijn	Eindhoven University of Technology, The Netherlands
Sašo Džeroski	Jožef Stefan Institute, Slovenia
Johannes Fürnkranz	Johannes Kepler University Linz, Austria
Mohamed Gaber	Birmingham City University, UK
Dragan Gamberger	Rudjer Bošković Institute, Croatia
Manco Giuseppe	Italian National Research Council, Italy
Kouichi Hirata	Kyushu Institute of Technology, Japan
Dino Ienco	Irstea, France
Nathalie Japkowicz	American University, USA
Stefan Kramer	Johannes Gutenberg University Mainz, Germany
Vincenzo Lagani	Ilija State University, Georgia
Pedro Larranaga	University of Madrid, Spain
Nada Lavrač	Jožef Stefan Institute, Slovenia
Tomislav Lipic	Rudjer Bošković Institute, Croatia
Gjorgji Madjarov	Ss. Cyril and Methodius University in Skopje, North Macedonia
Rafael Gomes Mantovani	Federal University of Technology – Paraná, Brazil
Elio Masciari	Institute for High Performance Computing and Networking, Italy
Nuno Moniz	University of Porto, Portugal
Anna Monreale	University of Pisa, Italy

Catarina Oliveira	Portucalense University, Portugal
Sageev Oore	Dalhousie University and Vector Institute, Canada
Rita P. Ribeiro	University of Porto, Portugal
George Papakostas	International Hellenic University, Greece
Ruggero G. Pensa	University of Turin, Italy
Pedro Pereira Rodrigues	University of Porto, Portugal
Bernhard Pfahringer	University of Waikato, New Zealand
Gianvito Pio	University of Bari Aldo Moro, Italy
Pascal Poncelet	LIRMM Montpellier, France
Jan Ramon	Inria, France
André L. D. Rossi	Universidade Estadual Paulista, Brazil
Kazumi Saito	University of Shizuoka, Japan
Tomislav Smuc	Rudjer Bošković Institute, Croatia
Marina Sokolova	University of Ottawa and Dalhousie University, Canada
Jerzy Stefanowski	Poznan University of Technology, Poland
Luis Filipe Teixeira	University of Porto, Portugal
Herna Viktor	University of Ottawa, Canada
Baghoussi Yassine	University of Porto, Portugal
Albrecht Zimmermann	University of Caen Normandy, France

Additional Reviewers

Ana Mestrovic	João Alves
Daniela Santos	Leonardo Ferreira
Emanuel Guberovic	Miha Keber
Fabrizio Lo Scudo	Narjes Davari
Francesca Pratesi	Nima Dehmamy
Francesco S. Pisani	Nunziato Cassavia
Francesco Spinnato	Paolo Mignone
Isabel Rio-Torto	Rafaela Leal
Ivan Grubisic	Roberto Interdonato
João Almeida	

Contents

Applications

Automated Grading of Exam Responses: An Extensive Classification Benchmark	3
<i>Jimmy Ljungman, Vanessa Lisevand, John Pavlopoulos, Alexandra Farazouli, Zed Lee, Panagiotis Papapetrou, and Uno Fors</i>	
Automatic Human-Like Detection of Code Smells	19
<i>Chitsutha Soomlek, Jan N. van Rijn, and Marcello M. Bonsangue</i>	
HTML-LSTM: Information Extraction from HTML Tables in Web Pages Using Tree-Structured LSTM	29
<i>Kazuki Kawamura and Akihiro Yamamoto</i>	
Predicting Reach to Find Persuadable Customers: Improving Uplift Models for Churn Prevention	44
<i>Théo Verhelst, Jeevan Shrestha, Denis Mercier, Jean-Christophe Dewitte, and Gianluca Bontempi</i>	

Classification

A Semi-supervised Framework for Misinformation Detection	57
<i>Yueyang Liu, Zois Boukouvalas, and Nathalie Japkowicz</i>	
An Analysis of Performance Metrics for Imbalanced Classification	67
<i>Jean-Gabriel Gaudreault, Paula Branco, and João Gama</i>	
Combining Predictions Under Uncertainty: The Case of Random Decision Trees	78
<i>Florian Busch, Moritz Kulessa, Eneldo Loza Mencía, and Hendrik Blockeel</i>	
Shapley-Value Data Valuation for Semi-supervised Learning	94
<i>Christie Courtage and Evgueni Smirnov</i>	

Data Streams

A Network Intrusion Detection System for Concept Drifting Network Traffic Data	111
<i>Giuseppina Andresini, Annalisa Appice, Corrado Loglisci, Vincenzo Belvedere, Domenico Redavid, and Donato Malerba</i>	

Incremental k -Nearest Neighbors Using Reservoir Sampling for Data Streams	122
<i>Maroua Bahri and Albert Bifet</i>	
Statistical Analysis of Pairwise Connectivity	138
<i>Georg Kreml, Daniel Kottke, and Tuan Pham</i>	
Graph and Network Mining	
FHA: Fast Heuristic Attack Against Graph Convolutional Networks	151
<i>Haoxi Zhan and Xiaobing Pei</i>	
Ranking Structured Objects with Graph Neural Networks	166
<i>Clemens Damke and Eyke Hüllermeier</i>	
Machine Learning for COVID-19	
Knowledge Discovery of the Delays Experienced in Reporting COVID-19 Confirmed Positive Cases Using Time to Event Models	183
<i>Aleksandar Novakovic, Adele H. Marshall, and Carolyn McGregor</i>	
Multi-scale Sentiment Analysis of Location-Enriched COVID-19 Arabic Social Data	194
<i>Tarek Elsaka, Imad Afyouni, Ibrahim Abaker Targio Hashem, and Zaher AL-Aghbari</i>	
Prioritization of COVID-19-Related Literature via Unsupervised Keyphrase Extraction and Document Representation Learning	204
<i>Blaž Škrlić, Marko Jukić, Nika Eržen, Senja Pollak, and Nada Lavrač</i>	
Sentiment Nowcasting During the COVID-19 Pandemic	218
<i>Ioanna Miliou, John Pavlopoulos, and Panagiotis Papapetrou</i>	
Neural Networks and Deep Learning	
A Sentence-Level Hierarchical BERT Model for Document Classification with Limited Labelled Data	231
<i>Jinghui Lu, Maeve Henchion, Ivan Bacher, and Brian Mac Namee</i>	
Calibrated Resampling for Imbalanced and Long-Tails in Deep Learning	242
<i>Colin Bellinger, Roberto Corizzo, and Nathalie Japkowicz</i>	
Consensus Based Vertically Partitioned Multi-layer Perceptrons for Edge Computing	253
<i>Haimonti Dutta, Saurabh Amarnath Mahindre, and Nitin Nataraj</i>	

Controlling BigGAN Image Generation with a Segmentation Network	268
<i>Aman Jaiswal, Harpreet Singh Sodhi, Mohamed Muzamil H, Rajveen Singh Chandhok, Sageev Oore, and Chandramouli Shama Sastry</i>	
GANs for Tabular Healthcare Data Generation: A Review on Utility and Privacy	282
<i>João Coutinho-Almeida, Pedro Pereira Rodrigues, and Ricardo João Cruz-Correia</i>	
Preferences and Recommender Systems	
An Ensemble Hypergraph Learning Framework for Recommendation	295
<i>Alireza Gharahighehi, Celine Vens, and Konstantinos Pliakos</i>	
KATRec: Knowledge Aware aTtentive Sequential Recommendations	305
<i>Mehrnaz Amjadi, Seyed Danial Mohseni Taheri, and Theja Tulabandhula</i>	
Representation Learning and Feature Selection	
Elliptical Ordinal Embedding	323
<i>Aïssatou Diallo and Johannes Fürnkranz</i>	
Unsupervised Feature Ranking via Attribute Networks	334
<i>Urh Primožič, Blaž Škrlj, Sašo Džeroski, and Matej Petković</i>	
Responsible Artificial Intelligence	
Deriving a Single Interpretable Model by Merging Tree-Based Classifiers	347
<i>Valerio Bonsignori, Riccardo Guidotti, and Anna Monreale</i>	
Ensemble of Counterfactual Explainers	358
<i>Riccardo Guidotti and Salvatore Ruggieri</i>	
Learning Time Series Counterfactuals via Latent Space Representations	369
<i>Zhendong Wang, Isak Samsten, Rami Mochaourab, and Panagiotis Papapetrou</i>	
Leveraging Grad-CAM to Improve the Accuracy of Network Intrusion Detection Systems	385
<i>Francesco Paolo Coforio, Giuseppina Andresini, Gennaro Vessio, Annalisa Appice, and Donato Malerba</i>	
Local Interpretable Classifier Explanations with Self-generated Semantic Features	401
<i>Fabrizio Angiulli, Fabio Fassetti, and Simona Nisticò</i>	

Privacy Risk Assessment of Individual Psychometric Profiles	411
<i>Giacomo Mariani, Anna Monreale, and Francesca Naretto</i>	
The Case for Latent Variable Vs Deep Learning Methods in Misinformation Detection: An Application to COVID-19	422
<i>Caitlin Moroney, Evan Crothers, Sudip Mittal, Anupam Joshi, Tülay Adalı, Christine Mallinson, Nathalie Japkowicz, and Zois Boukouvalas</i>	
Spatial, Temporal and Spatiotemporal Data	
Local Exceptionality Detection in Time Series Using Subgroup Discovery: An Approach Exemplified on Team Interaction Data	435
<i>Dan Hudson, Travis J. Wiltshire, and Martin Atzmueller</i>	
Neural Additive Vector Autoregression Models for Causal Discovery in Time Series	446
<i>Bart Bussmann, Jannes Nys, and Steven Latré</i>	
Spatially-Aware Autoencoders for Detecting Contextual Anomalies in Geo-Distributed Data	461
<i>Roberto Corizzo, Michelangelo Ceci, Gianvito Pio, Paolo Mignone, and Nathalie Japkowicz</i>	
Author Index	473