Lecture Notes in Artificial Intelligence 12323

Subseries of Lecture Notes in Computer Science

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

Annalisa Appice · Grigorios Tsoumakas ·
Yannis Manolopoulos · Stan Matwin (Eds.)

Discovery Science

23rd International Conference, DS 2020
Thessaloniki, Greece, October 19–21, 2020
Proceedings

Springer

Editors
Annalisa Appice
University of Bari Aldo Moro
Bari, Italy

Yannis Manolopoulos
Open University of Cyprus
Nicosia, Cyprus

Grigorios Tsoumakas
Aristotle University of Thessaloniki
Thessaloniki, Greece

Stan Matwin
Dalhousie University
Halifax, NS, Canada

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-030-61526-0 ISBN 978-3-030-61527-7 (eBook)
https://doi.org/10.1007/978-3-030-61527-7

LNCS Sublibrary: SL7 – Artificial Intelligence

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 23rd International Conference on Discovery Science (DS 2020), which was organized to be held in Thessaloniki, Greece, during October 19–21, 2020. Due to the outbreak of the COVID-19 pandemic, the conference was moved online and held virtually over the same time period. The conference was organized by the Aristotle University of Thessaloniki, Greece, in cooperation with the Open University of Cyprus, Cyprus, Dalhousie University, Canada, and University of Bari Aldo Moro, Italy.

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, with their application in various scientific domains. In particular, major areas selected for DS 2020 include: artificial intelligence applied to science; machine learning; knowledge discovery and data mining; causal modeling; AutoML, meta-learning, and planning to learn; machine learning and high-performance computing; grid and cloud computing; literature-based discovery; ontologies for science, including the representation and annotation of datasets and domain knowledge; explainable AI, interpretability of machine learning, and deep learning models; process discovery and analysis; computational creativity; anomaly detection and outlier detection; data streams, evolving data, change detection, concept drift, and model maintenance; network analysis; time-series analysis; learning from complex data; data and knowledge visualization; human-machine interaction for knowledge discovery and management; evaluation of models and predictions in discovery setting; machine learning and cybersecurity; as well as applications of the above techniques in scientific domains.

DS 2020 received 76 international submissions that were carefully reviewed by three or more Program Committee (PC) members or external reviewers. After a rigorous reviewing process, 26 regular papers and 19 short papers were accepted for presentation at the conference and publication in the DS 2020 volume. Short papers were just allotted a smaller presentation time compared to regular ones.

The conference program included three invited keynotes. Prof. Myra Spiliopoulou from Otto von Guericke University Magdeburg, Germany contributed a talk titled "Knowledge Discovery in mHealth – dealing with few noisy data." Prof. Peter A. Flach from University of Bristol, UK, contributed a talk titled "The highs and lows of performance evaluation: Towards a measurement theory for machine learning." Prof. Gustau Camps-Valls from Universitat de València, Spain, gave a presentation titled "Machine learning for Modelling and Understanding in Earth Sciences." Abstracts of the invited talks with short biographies of the invited speakers are included in this volume.

We would like to sincerely thank all people who helped this volume come into being and made DS 2020 a successful and exciting event. In particular, we would like to

express our appreciation for the work of the DS 2020 PC members and external reviewers who helped assure the high standard of accepted papers. We would like to thank all authors of DS 2020, without whose high-quality contributions it would not have been possible to organize the conference.

We are grateful to the Steering Committee chair, Sašo Džeroski, and the whole Steering Committee for their extraordinary support in critical decisions concerning the event plan. We wish to express our thanks to local organization chairs, Anastasios Gounaris and Apostolos Papadopoulos, and the whole organization team for their support and incredible work. We would also thank the treasurer, Richard Chbeir, for his professional 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 2020 for a chance to be published in this prestigious journal. For DS 2020, Springer also supported a Best Paper Award to Riku Laine, Antti Hyttinen, and Michael Mathioudakis for their paper "Evaluating Decision Makers over Selectively Labelled Data: A Causal Modeling Approach." We would also like to honorary mention the runner-up paper "Explaining Sentiment Classification with Synthetic Exemplars and Counter-Exemplars" by Orestis Lampridis, Riccardo Guidotti, and Salvatore Ruggieri.

September 2020

Annalisa Appice
Grigorios Tsoumakas
Yannis Manolopoulos
Stan Matwin

Organization

General Chairs

Yannis Manolopoulos	Open University of Cyprus, Cyprus
Stan Matwin	Dalhousie University, Canada

Program Committee Chairs

Annalisa Appice	University of Bari Aldo Moro, Italy
Grigorios Tsoumakas	Aristotle University of Thessaloniki, Greece

Program Committee

Martin Atzmüller	Tilburg University, The Netherlands
Viktor Bengs	Paderborn University, Germany
Concha Bielza Lozoya	Universidad Politécnica de Madrid, Spain
Konstantinos Blekas	University of Ioannina, Greece
Alberto Cano	Virginia Commonwealth University, USA
Michelangelo Ceci	University of Bari Aldo Moro, Italy
Paolo Ceravolo	Politecnico di Milano, Italy
Bruno Cremilleux	University of Caen Normandy, France
Claudia d'Amato	University of Bari Aldo Moro, Italy
Nicola Di Mauro	University of Bari Aldo Moro, Italy
Ivica Dimitrovski	Ss. Cyril and Methodius University, North Macedonia
Wouter Duivesteijn	Eindhoven University of Technology, The Netherlands
Sašo Džeroski	Jožef Stefan Institute, Slovenia
Hadi Fanaee-T	University of Oslo, Norway
Nicola Fanizzi	University of Bari Aldo Moro, Italy
Stefano Ferilli	University of Bari Aldo Moro, Italy
Johannes Fürnkranz	Johannes Kepler University Linz, Austria
Mohamed Gaber	Birmingham City University, UK
Dragan Gamberger	Rudjer Bošković Institute, Croatia
Dimitris Gunopoulos	University of Athens, Greece
Makoto Haraguchi	Hokkaido University, Japan
Kouichi Hirata	Kyushu Institute of Technology, Japan
Jaakko Hollmén	Aalto University, Finland
Eyke Hüllermeier	Paderborn University, Germany
Dino Ienco	Irstea, France
Alípio Jorge	University of Porto, Portugal
Ioannis Katakis	University of Nicosia, Cyprus
Masahiro Kimura	Ryukoku University, Japan
Dragi Kocev	Jožef Stefan Institute, Slovenia

Petra Kralj Novak	Jožef Stefan Institute, Slovenia
Stefan Kramer	Johannes Gutenberg University Mainz, Germany
Vincenzo Lagani	Ilia State University, USA
Pedro Larranaga	University of Madrid, Spain
Nada Lavrač	Jožef Stefan Institute, Slovenia
Jurica Levatić	Institute for Research in Biomedicine, Spain
Tomislav Lipic	Rudjer Bošković Institute, Croatia
Francesca A. Lisi	University of Bari Aldo Moro, Italy
Gjorgji Madjarov	Ss. Cyril and Methodius University, North Macedonia
Giuseppe Manco	Institute for High Performance Computing and Networking, Italy
Elio Masciari	Institute for High Performance Computing and Networking, Italy
Anna Monreale	University of Pisa, Italy
Rita P. Ribeiro	University of Porto, Portugal
Panče Panov	Jožef Stefan Institute, Slovenia
George Papakostas	International Hellenic University, Greece
Ruggero G. Pensa	University of Torino, Italy
Bernhard Pfahringer	University of Waikato, New Zealand
Gianvito Pio	University of Bari Aldo Moro, Italy
Pascal Poncelet	LIRMM Montpellier, France
Chedy Raïssi	French Research Institute for Digital Sciences, France
Jan Ramon	French Research Institute for Digital Sciences, France
Kazumi Saito	University of Shizuoka, Japan
Tomislav Šmuc	Rudjer Bošković Institute, Croatia
Jerzy Stefanowski	Poznan University of Technology, Poland
Ljupčo Todorovski	University of Ljubljana, Slovenia
Luis Torgo	Dalhousie University, Canada
Ioannis Tsamardinos	University of Crete, Greece
Herna Viktor	University of Ottawa, Canada
Michalis Vlachos	Université de Lausanne, Switzerland
Albrecht Zimmermann	University of Caen Normandy, France
Blaž Zupan	University of Ljubljana, Slovenia

Additional Reviewers

Besher Alhalabi
Dridi Amna
Giuseppina Andresini
Emanuele Pio Barracchia
Martin Breskvar
Lorraine Chambers
Graziella De Martino
Massimo Guarascio
Julian Hatwell
Theofanis Kalampokas
Ana Kostovska
Ilona Kulikosvkikh
Masahito Kumano
Vladimir Kuzmanovski
Athanasios Lagopoulos
Paolo Mignone
John Mollas
Igor Mozetič
Vu-Linh Nguyen
Luca Oneto
Vincenzo Pasquadibisceglie
Francesco Scicchitano
Tomaž Stepišnik
Bozhidar Stevanoski
Alexander Tornede
Aleš Žagar

Abstracts of Keynote Talks

Knowledge Discovery in mHealth – Dealing with Few Noisy Data

Myra Spiliopoulou

Research Group on Knowledge Management and Discovery (KMD), Faculty of Computer Science, Otto von Guericke University Magdeburg, PO Box 4120, 39016 Magdeburg, Germany
myra@iti.cs.uni-magdeburg.de

Abstract. Patients with chronic diseases can greatly benefit from mHealth technology. There are solutions assisting them in measuring signals (e.g., blood pressure, sugar level, etc.), in keeping a diary with Ecological Momentary Assessments (EMA), such as physical exercise, onset of symptoms, and subjective perception of health condition. Machine learning can deliver useful insights from data thus collected. While sensor signals can be collected without interruption, EMA recording depends on patients' self-discipline and compliance.

The talk starts with an overview of the role of mHealth applications in diagnostics and treatment support. Then, we focus on EMA for chronic conditions. We discuss challenges of learning from few and noisy recordings, and methods for prediction and risk factor identification on these data.

Keywords: mHealth · Multidimensional sequences · Gaps · Time series prediction · Adherence

The Highs and Lows of Performance Evaluation: Towards a Measurement Theory for Machine Learning

Peter A. Flach

Intelligent Systems Laboratory, Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, UK
Peter.Flach@bristol.ac.uk

Abstract. Our understanding of performance evaluation measures for machine-learned classifiers has improved considerably over the last decades. However, there is a range of areas where this understanding is still lacking, leading to ill-advised practices in classifier evaluation. This is clearly problematic, since if machine learning researchers are unclear about what exactly their experiments are telling them about their machine learning algorithms, then how can end-users trust systems deploying those algorithms?

I suggest that in order to make further progress we need to develop a proper measurement theory of machine learning. Measurement theory studies the concepts of measurement and scale. If you have a way to measure, say, the length of individual rods or planks, this should also allow you to then calculate the combined length of concatenated rods or planks. What relevant concatenation operations are there in data science and AI, and what does that mean for the underlying measurement scale?

I discuss by example what such a measurement theory might look like and what kinds of new results it would entail. I furthermore argue that key properties such as classification ability and data set difficulty are unlikely to be directly observable, suggesting the need for latent-variable models. Ultimately, machine learning experiments need to go beyond simple correlations and aim to make causal inferences of the form 'Algorithm A outperformed algorithm B because two classes were highly imbalanced,' or counterfactually, 'if the classes were rebalanced, the observed performance difference between A and B would disappear.'

Keywords: Machine learning experiments · Classification performance · Psychometrics · latent variables · Levels of measurement · Causal inference

Machine Learning for Modelling and Understanding in Earth Sciences

Gustau Camps-Valls

Image Processing Lab, Universitat de València, Spain
https://isp.uv.es
@isp_uv_es

Abstract. The Earth is a complex dynamic network system. Modelling and understanding the system is at the core of scientific endeavour. We approach these problems with machine learning (ML) algorithms. I will review several ML approaches we have developed in the last years: 1) advanced Gaussian processes models for bio-geo-physical parameter estimation, which can incorporate physical laws, blend multisensor data while providing credible confidence intervals for the estimates, and improved interpretability, 2) nonlinear dimensionality reduction methods to decompose Earth data cubes in spatially-explicit and temporally-resolved modes of variability that summarize the information content of the data and allow for identifying relations with physical processes, and 3) advances in causal inference that can uncover cause and effect relations from purely observational data.

Contents

Ensembles

Explainable and Interpretable Machine Learning

Graph and Network Mining

Multi-target Models

Neural Networks and Deep Learning

Spatial, Temporal and Spatiotemporal Data