Lecture Notes in Computer Science

Lecture Notes in Artificial Intelligence 15244

Founding Editor

Jörg Siekmann

Series Editors

The series Lecture Notes in Artificial Intelligence (LNAI) was established in 1988 as a topical subseries of LNCS devoted to artificial intelligence.

The series publishes state-of-the-art research results at a high level. As with the LNCS mother series, the mission of the series is to serve the international R & D community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings.

Dino Pedreschi · Anna Monreale ·
Riccardo Guidotti · Roberto Pellungrini ·
Francesca Naretto
Editors

Discovery Science

27th International Conference, DS 2024
Pisa, Italy, October 14–16, 2024
Proceedings, Part II

Springer

Editors
Dino Pedreschi
University of Pisa
Pisa, Italy

Anna Monreale
University of Pisa
Pisa, Italy

Riccardo Guidotti
University of Pisa
Pisa, Pisa, Italy

Roberto Pellungrini
Scuola Normale Superiore (SNS)
Pisa, Italy

Francesca Naretto
University of Pisa
Pisa, Italy

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-031-78979-3 ISBN 978-3-031-78980-9 (eBook)
https://doi.org/10.1007/978-3-031-78980-9

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

Discovery Science 2024 conference provides an open forum for intensive discussions and exchange of new ideas among researchers working in the area of Discovery Science. The conference focus is on the use of Artificial Intelligence, Data Science and Big Data Analytics methods in science. Its scope includes the development and analysis of methods for discovering scientific knowledge, coming from machine learning, data mining, intelligent data analysis, and big data analytics, as well as their application in various domains. The 27th International Conference on Discovery Science (DS 2024) was held in Pisa, Italy, during October 14–16, 2024.

This was the fourth time the conference was organized as a stand-alone physical event. Indeed, for its first 20 editions, DS was co-located with the International Conference on Algorithmic Learning Theory (ALT). In 2018 it was co-located with the 24th International Symposium on Methodologies for Intelligent Systems (ISMIS 2018). Then starting from 2019, it has been a stand-alone event. DS 2020 and DS 2021 were online-only events while DS 2022 and DS 2023 were located in Montpellier, France and Porto, Portugal, respectively.

DS 2024 received 121 international submissions, of which 25 were for the SoBigData++ track, a track dedicated to the usage of data and data science in science celebrating the conclusion of the SoBigData++ project, a sponsor of DS 2024. For each track, each submission was reviewed by at least two Program Committee (PC) members (some PC acted for both tracks) in a single-blind review process using the Microsoft CMT system. The PC decided to accept 45 papers for the regular research track of DS 2024 and 9 papers for the SoBigData++ track, for a total of 54 papers. This resulted in an overall acceptance rate of 45%. One paper was withdrawn two weeks before the beginning of the conference. Thus, during the conference, 53 papers were presented and included in these volumes.

The conference included three keynote talks: Roberto Navigli (Sapienza University of Rome and Babelscape) contributed a talk titled "What Is Missing in Today's Large Language Models?"; Carlos Castillo (ICREA and Universitat Pompeu Fabra) gave a presentation titled "Human Factors and Algorithmic Fairness"; and Francesca Toni (Imperial College London) contributed a talk titled "Bridging Explainable AI and Contestability". Abstracts of the invited talks are included in the front matter of these proceedings. Besides the presentation of the regular research papers and SoBigData++ papers in the main program, the conference offered two poster sessions titled "Late Breaking Contributions" and "Doctoral Consortium", featuring posters of very recent research results and PhD theses on topics related to Discovery Science.

We are grateful to Springer for their continued long-term support. Springer publishes the conference proceedings, as well as a regular special issue of the Machine Learning journal on Discovery Science. The latter offers authors a chance of publishing in this prestigious journal significantly extended and reworked versions of their DS conference papers, while being open to all submissions on DS conference topics.

On the program side, we would like to thank all the authors of the submitted papers and the PC members for their efforts in evaluating the submitted papers, as well as the keynote speakers. On the organization side, we would like to thank all the members of the Organizing Committee, in particular Roberto Pellungrini, Francesca Naretto, Vittorio Romano, Francesco Spinnato, Lorenzo Mannocci, Daniele Fadda and Rosalba Lubino for the smooth preparation and organization of all conference-associated activities. We are also grateful to the people behind Microsoft CMT for developing the conference organization system which has proved to be an essential tool in the paper submission and evaluation process.

October 2024

Dino Pedreschi
Anna Monreale
Riccardo Guidotti
Roberto Pellungrini
Francesca Naretto

Organization

General and Program Chairs

Dino Pedreschi	University of Pisa, Italy
Anna Monreale	University of Pisa, Italy
Riccardo Guidotti	University of Pisa, Italy

Special Session Chair

Roberto Trasarti	ISTI-CNR Pisa, Italy

Poster Session Chair

Francesca Naretto	University of Pisa, Italy

Doctoral Consortium Chair

Roberto Pellungrini	Scuola Normale Superiore, Italy
Fosca Giannotti	Scuola Normale Superiore, Italy

Social Media and Publicity Chair

Vittorio Romano	ISTI-CNR Pisa, Italy

Local Organization Committee

Francesco Spinnato	University of Pisa, Italy
Lorenzo Mannocci	University of Pisa, Italy
Daniele Fadda	ISTI-CNR Pisa, Italy

Steering Committee Chair

Michelangelo Ceci University of Bari, Italy

Program Committee

Ad Feelders
Adriano Rivolli
Alberto Cano
Albrecht Zimmermann
Alexander H. Gower
Andreas Nuernberger
Andrew Aquilina
Angelica Liguori
Annalisa Appice
Anne Laurent
Antonio Mastropietro
Apostolos N. Papadopoulos
Arnaud Soulet
Bernard Zenko
Bernhard Pfahringer
Blaz Zupan
Brian Mac Namee
Bruno Cremilleux
Bruno E. Martins
Bruno Veloso
Carlo Metta
Carlos Soares
Caterina Senette
Chamalee Nisanala Wickrama Arachchi
Chetraj Pandey
Chiara Renso
Claire Nédellec
Clara R. T. Puga
Claudio Giovannoni
Cristiano Landi
Daniele Gambetta
Dariusz Brzezinski
Davide Vega
Dino Ienco
Domen Šoberl
Domenico Talia
Donato Malerba
Dragan Gamberger
Dragi Kocev
Elio Masciari
Eric Scott
Esther-Lydia Silva-Ramírez
Fabien FP Poirier
Fabio Fassetti
Fabrizio Angiulli
Fabrizio Marozzo
Federico Mazzoni
Florent Masseglia
Francesca Bugiotti
Francesca Naretto
Francesca Alessandra Lisi
Francesco Marcelloni
Francesco S. Pisani
Francesco Spinnato
Georgios Vardakas
Gianvito Pio
Giuseppe Manco
Giuseppina Andresini
Gizem Gezici
Günce Keziban Orman
Gustau Camps-Valls
Haimonti Dutta
Henrik Bostrom
Hoang-Anh Ngo
Howard J. Hamilton
Inês Dutra
Isacco Beretta
Ivan Reis Filho
Jaka Kokošar
Jelena Joksimovic
Joana Santos
Joao Mendes-Moreira
Johannes Fürnkranz
Jörg Wicker

José Luis Seixas Jr.
Juan G. Colonna
Julian Martin Rodemann
Julian Marvin Joers
Julio C. Muñoz-Benítez
Kai Puolamäki
Katharina Dost
Kaustubh Patil
Kohei Hatano
Larisa Soldatova
Larissa Andrade Silva
Lorenzo Mannocci
Lubos Popelínský
Luca Corbucci
Luca Ferragina
Maguelonne Teisseire
Maik Büttner
Marcilio de Souto
Marco Javier Suárez Barón
Marek Herde
Margarida A. Costa
Marina Sokolova
Mário Antunes
Marko Pranjifá
Marta Marchiori Manerba
Marta Moreno
Martin Atzmueller
Martin Holena
Martin Špendl
Martina Cinquini
Masaaki Nishino
Massimo Guarascio
Mathieu Roche
Matteo Zignani
Mattia Cerrato
Mattia Setzu
Maurizio Parton
Maximilien Servajean
Michael R. Berthold
Michele Fontana
Mirco Nanni
Myra Spiliopoulou
Nada Lavrafç
Nathalie Japkowicz
Nontokozo Mpofu
Nuno Silva
Panagiotis Papapetrou
Pance Panov
Pascal Poncelet
Paula Silva
Paulo Cortez
Pavel Brazdil
Pavlin G. Polifçar
Pedro C. Vieira
Pedro G. Ferreira
Pedro Henriques Abreu
Peter O. Koleoso
Peter van der Putten
Rafael Gomes Mantovani
Rafael Mamede
Rafael G. Teixeira
Raza Ul Mustafa
Reza Akbarinia
Ricardo Cardoso Pereira
Ricardo Cerri
Riccardo Cantini
Rita D. Nogueira
Robert Bossy
Roberto Corizzo
Roberto Interdonato
Roberto Pellungrini
Ross King
Rui Jorge Gomes
Sabrina Gaito
Salvatore Ruggieri
Saso Dzeroski
Silvia Corbara
Simona Nisticò
Simone Piaggesi
Sónia Teixeira
Takayasu Fushimi
Thiago Andrade
Van Anh Huynh-Thu
Vincent Labatut
Vincenzo Lagani
Vincenzo Pasquadibisceglie
Xuan Zhao
Yiping Tang
Zahra Donyavi
Zakaria Farou

Program Committee SoBigData++ Track

Albert Ali Salah
Antinisca Di Marco
Chiara Boldrini
Clara Punzi
Claudio Giovannoni
Fabrizia Auletta
Fabrizio Lillo
George Papastefanatos
Giovanni Stilo
Josep Domingo-Ferrer
Mark E. Cote
Marzio Di Vece
Maurizio Parton
Michela Natilli
Paolo Bellavista
Rami Haffar
Roberto Trasarti
Ruggero G. Pensa
Tom Emery
Valerio Grossi

Keynote Talks

What Is Missing in Today's Large Language Models?

Roberto Navigli

Sapienza University of Rome and Babelscape, Italy

Large Language Models (LLMs) like GPT-4 have demonstrated remarkable capabilities in generating human-like text, understanding context, and performing a wide range of tasks across various domains. However, despite their impressive performance, LLMs exhibit critical limitations that restrict their applicability and reliability in real-world scenarios. This talk delves into the key areas where LLMs fall short, including the lack of true understanding and reasoning, susceptibility to biases, difficulties with long-term context retention, and challenges in generating accurate outputs in non-predominant domains. I will also touch upon research directions that can provide potential solutions to enhance the capabilities and trustworthiness of future LLMs, such as performing lexical and sentence-level semantics, intersecting knowledge and results obtained from different tasks, improving training data diversity, and supporting factuality.

Human Factors and Algorithmic Fairness

Carlos Castillo

ICREA and Universitat Pompeu Fabra, Spain

In this talk, we present ongoing research on human factors of decision support systems that has consequences from the perspective of algorithmic fairness. We study two different settings: a game and a high-stakes scenario. The game is an exploratory "oil drilling" game, while the high-stakes scenario is the prediction of criminal recidivism. In both cases, a decision support system helps a human make a decision. We observe that in general users of such systems must thread a fine line between algorithmic aversion (completely disregarding the algorithmic support) and automation bias (completely disregarding their own judgment). The talk presents joint work led by David Solans and Manuel Portela.

Bridging Explainable AI and Contestability

Francesca Toni

Imperial College London, UK

AI has become pervasive in recent years, and the need for explainability is widely agreed upon as crucial towards safe and trustworthy deployment of AI systems. However, state-of-the-art AI and eXplainable AI (XAI) approaches mostly neglect the need for AI systems to be contestable, as advocated instead by AI guidelines (e.g. by the OECD) and regulation of automated decision-making (e.g. GDPR in the EU and UK). In this talk I will advocate forms of contestable AI that can (1) interact to progressively explain outputs and/or reasoning, (2) assess grounds for contestation provided by humans and/or other machines, and (3) revise decision-making processes to redress any issues successfully raised during contestation. I will then explore how contestability can be achieved computationally, starting from various approaches to explainability, including some drawn from the field of computational argumentation. Specifically, I will overview a number of approaches to (argumentation-based) XAI for neural models and for causal discovery and their uses to achieve contestability.

Contents – Part II

Computer Vision and Explainable AI

Classification Models

SoBigData++: City for Citizens and Explainable AI