3rd Workshop on Learning with Imbalanced Domains: Preface

Nuno Moniz
INESC TEC / Faculty of Sciences, University of Porto
Porto, Portugal
nmmoniz@inesctec.pt

Paula Branco
School of Electrical Engineering and Computer Science, University of Ottawa
Ontario, Canada
pbranco@uottawa.ca

Luís Torgo
Faculty of Computer Science, Dalhousie University
Halifax, Canada
ltorgo@dal.ca

Nathalie Japkowicz
Department of Computer Science, American University
Washington DC, USA
japkowic@american.edu

Michał Woźniak
Wroclaw University of Science and Technology
Wrocław, Poland
michal.wozniak@pwr.edu.pl

Shuo Wang
University of Birmingham
Birmingham, UK
s.wang.2@bham.ac.uk

This volume contains the Proceedings of the Third International Workshop on Learning with Imbalanced Domains: Theory and Applications - LIDTA 2021. This Workshop was co-organised by INESC TEC, the Department of Computer Science at the Faculty of Sciences of the University of Porto (Portugal), the School of Electrical Engineering and Computer Science at the University of Ottawa (Canada), the Faculty of Computer Science at Dalhousie University (Canada), the Department of Computer Science of the American University (USA), the Wroclaw University of Science and Technology (Poland) and the University of Birmingham (UK). The Workshop was co-located with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD) 2021 and was held as an online event on the 17th of September 2021.

The LIDTA 2021 Workshop focused on both theoretical and practical aspects of the problem of learning from imbalanced domains. In multiple real-world applications, the end-user aims at obtaining predictive models that are able to reflect her/his domain preferences. When these preferences are not uniform over the target variable domain, this causes a problem. Non-uniform preferences are critical in imbalanced domains where we observe that the most relevant target variable values for the end-user are scarcely represented. This problem is evident in many real-world domains such as financial (Kamalov, 2020), medical (Cao et al., 2018), meteorological (Troncoso et al., 2018), cybersecurity (Wheelus et al., 2018) or social media (Li and Liu, 2018).
The problem of imbalanced domains has been extensively studied in the last decade for binary classification tasks. The study of this problem in other predictive contexts has been gaining more attention in the recent years. In fact, several researchers are now focused on tackling this problem in the context of multiclass problems (Koziarski et al., 2020), regression tasks (Torgo et al., 2013), multi-label classification (Charte et al., 2019), association rules mining (Luna et al., 2015), multi-instance learning (Vluymans et al., 2016), data streams (Krawczyk et al., 2017), and time series (Moniz et al., 2017), among others. This is in fact a broad issue involving multiple challenges which is common to a diversity of tasks.

The LIDTA 2021 workshop is focused on these problems. Following the trend of the previous editions of this workshop, LIDTA 2021 received a diverse set of contributions. The selected papers are high quality, inter-disciplinary articles that discuss numerous aspects of the problem of learning from imbalanced domains. Overall, there were 13 paper submissions, out of which 8 papers were accepted for inclusion in the workshop proceedings. The high workshop attendance reflected the great interest of the research community in the topic. The workshop included a morning session and an afternoon session. Each session included a keynote talk. After the welcoming, the morning session started with an invited talk entitled “Learning with Imbalanced Data Streams”, by Professor Bartosz Krawczyk, from the Department of Computer Science, Virginia Commonwealth University. The afternoon session started with our second invited talk from Professor Nathalie Japkowicz from the Department of Computer Science, American University, Washington DC, USA, entitled “Class Imbalances and Deep Learning”. Both talks raised several interesting questions and remarks from the audience. The success of this workshop edition which builds on the previous workshops accomplishments enabled a follow-up Special Issue on Imbalanced Learning, hosted by the Machine Learning Journal.

All the papers accepted in LIDTA 2021 workshop were assigned a 15-minute presentation slot, followed by 5 minutes for questions and answers. Four papers were presented in each one of the morning and afternoon sessions. More details about the accepted papers are described next.

Nardi et al. (2021) address the anomaly detection problem in decentralized scenarios. The authors propose an unsupervised ensemble method for this problem in which the base learners are lightweight autoencoders. Sadeghi and Viktor (2021) present Online-MC-Queue algorithm, a novel solution for online learning in the context of multi-class imbalanced problems. The authors use a queue-based resampling method that creates an instance queue for each problem class. This queue maintains the number of instances balanced. A novel algorithm named Multi-label Neighbourhood Component Analysis (ML-NCA) is proposed by Pakrashi et al. (2021). The ML-NCA is designed for addressing issues on multi-label classification problems. ML-NCA performs a supervised linear transformation of the input space to obtain a new space where KNN-based algorithms are expected to perform well. Draghi et al. (2021) explore methods to improve synthetic data generators. The authors use probabilistic methods to identify difficult to predict data samples, and then use these methods to boost these types of data when generating synthetic samples. Limnios et al. (2021) present a new method for outlier detection in the presence of vast amounts of normal data. The authors propose the learning of a data-driven scoring function that reflects the degree of abnormality of the observations. Nazari and Branco (2021) present an analysis of the impact of using CGANs as an oversampling strategy as a method to
tackle the class imbalance and other data difficult factors. Naklicka and Stefanowski (2021) provide a contribution related to the extension of the BRACID rule-based classifier to a multi-class imbalanced scenario. Two solutions are proposed: the first uses BRACID in the OVO ensemble along with modifications of the prediction aggregation strategy while the second changes the rules induction for multiple classes simultaneously. Finally, Bougaham et al. (2021) present a solution for an application domain involving the use of intermediate patches, after a WGAN training procedure. The key goal of this approach is to enable the detection of anomalies on full size Printed Circuit Board Assembly (PCBA) images.

We would like to thank all of the authors and the Program Committee members for their hard work and commitment that allowed to carry out a successful and interesting workshop. We also want to deeply thank the ECML/PKDD 2021 Workshop and Tutorial Chairs for their support.

Organizing Committee

- Nuno Moniz (INESC TEC / Faculty of Sciences, University of Porto)
- Paula Branco (Faculty of Engineering, University of Ottawa)
- Luís Torgo (Faculty of Computer Science, Dalhousie University)
- Nathalie Japkowicz (Department of Computer Science, American University)
- Michał Woźniak (Wroclaw University of Science and Technology)
- Shuo Wang (University of Birmingham)

Program Committee

- Gustavo Batista, University of New South Wales
- Colin Bellinger, University of Alberta
- Seppe Vanden Broucke, Katholieke Universiteit Leuven
- Nitesh Chawla, University of Notre Dame
- Chris Drummond, NRC Institute for Information Technology
- Alberto Fernández, Granada University
- Mikel Galar, Universidad Pública de Navarra
- Salvador Garcia, University of Granada
- Raji Ghawi, Technical University of Munich
- Nikou Guennemann, Technical University of Munich
- Jose Hernandez-Orallo, Universitat Politecnica de Valencia
• Inaki Inza, University of the Basque Country
• Michal Koziarzki, AGH University of Science and Technology
• Bartosz Krawczyk, Virginia Commonwealth University
• Leandro Minku, University of Birmingham
• Ronaldo Prati, Universidade Federal do ABC - UFABC
• Rita Ribeiro, DCC - Faculty of Sciences, University of Porto
• Marina Sokolova, University of Ottawa
• Jerzy Stefanowski, Poznan University of Technology
• Herna Viktor, University of Ottawa
• Gary Weiss, Fordham University

References


