

Editorial: Fifth special issue on Knowledge Discovery and Business Intelligence

Paulo Cortez¹ Albert Bifet²

¹ ALGORITMI Research Centre, Department of Information
Systems, University of Minho, 4800-058 Guimarães, Portugal

Email: pcortez@dsi.uminho.pt

ORCID: 0000-0002-7991-2090

² Telecom ParisTech, Universite Paris-Saclay, 75634 Paris Cedex 13,
France

Email: albert.bifet@telecom-paristech.fr

ORCID: 0000-0002-8339-7773

1 Introduction

Artificial Intelligence (AI) is impacting our world. In the 1970s and 1980s, Expert Systems (ES) consisted of AI systems that included explicit knowledge, often represented in a symbolic form (e.g., by using the Prolog language), that was extracted from human experts. Since then, there has been an AI shift, due to three main phenomena [Darwiche, 2018]: data explosion, with availability of several big data sources (e.g., social media, sensor data); computational power growth, following the famous Moores law which states that computer processing capacity doubles every two years; and rise of sophisticated statistical and optimization techniques, including deep learning. Thus, rather than being expert-driven, ES have become more data-driven, with the focus on developing “computerized systems that use AI techniques to solve a specific real-world domain application task” [Cortez et al., 2018].

Aiming to foster the interaction between two key ES areas, Knowledge Discovery (KD) and Business Intelligence (BI), a series of “Knowledge Discovery and Business Intelligence” (KDBI) tracks were held at the EPIA conference on Artificial Intelligence, with a total of six editions from 2009 to 2019. Since 2011, the track has a dedicated special issue published in Wiley’s Expert Systems journal (EXSY) [Cortez and Santos, 2013, Cortez and Santos, 2015, Cortez and Santos, 2017, Cortez and Santos, 2018]. KD is the AI subfield that addresses the extraction of useful knowledge from raw data [Fayyad et al., 1996], while BI, also known as Business Analytics, is an umbrella term that includes methods and tools (e.g., data warehousing, dashboards and analytics) to obtain actionable knowledge from data [Ain et al., 2019].

This is the Fifth special issue on Knowledge Discovery and Business Intelligence and it includes extended versions of selected papers presented at the 6th KDBI

thematic track of EPIA 2019, held in Vila Real, Portugal. The track received a total of 17 paper submissions, from which 10 papers were accepted to be presented at the EPIA 2019 conference. The special issue of the EYSY journal involved two rounds of reviews for the selected papers, performed by the program committee members of the 6th KDBI track of EPIA2019 and EXSY journal expert reviewers. After the two rounds, six papers were accepted for the EXSY special issue.

The accepted KDBI special issue papers reflect current AI methodological and application challenges. For instance, nowadays image, sound, sensor and social media data are becoming commonplace, thus there is a need to develop machine learning systems capable of processing such data and proving value in real-world applications. Moreover, the data-driven models should be understandable by the domain humans, in what is termed as Explainable AI (XAI). Finally, the extracted data-driven knowledge should be actionable, allowing to better support managerial decision-making. These challenges are addressed by the six papers published in this special issue, which are summarized in the next section.

2 Contents of the special issue

The first paper, entitled ‘Semantic Segmentation and Colorization of Grayscale Aerial Imagery with W-Net Models’, involves the semantic segmentation of remotely sensed aerial images. Dias et al. (2020) propose the W-Net specialized deep learning architecture to simultaneously segment images and reconstruct the color of input images. Several experiments were held, with historical road maps and building footprints from the cities of Potsdam and Vaihingen, revealing competitive results when compared with the U-Net baseline model. Therefore, the adapted W-Net model can be used as a valuable image processing tool, alleviating the need for a manual inspection of aerial images.

Another deep learning model, the Convolutional Neural Network (CNN), was adapted by Anjos et al. (2020) to classify sibilant phonemes of European Portuguese sounds. The paper, entitled ‘Sibilant consonants classification comparison with multi and single-class neural networks’, proposes a serious sound game, aiming to help children to improve their speech. To benchmark the classifiers, a total of 1,500 sounds were collected using 145 children from three Portuguese schools. Mel-frequency Cepstral Coefficients (MFCCs) were used to preprocess the sounds, in order to generate the CNN inputs. The proposed CNN model outperformed two baseline models (a simpler Neural Network and a Support Vector Machine), obtaining a quality classification that can potentially help in language therapy games.

In ‘From Mobility Data to Habits and Common Pathways’, Andrade et al. (2020) analyze two real-world Global Positioning System (GPS) datasets, related with human trajectories from the cities of Beijing, China, and Porto, Portugal. In particular, a spatio-temporal clustering approach was proposed, in order to discover common pathways across the users daily habits. The authors have found that humans tend to follow a regular schedule and frequent routes when moving between their preferred locations. The devised models can be used to predict the next locations of users or groups of users, which can be valuable in several real-world applications, including smart cities.

The fourth paper, ‘A Context-Aware Recommender Method based on Text and Opinion Mining’, by Sundermann et al. (2020), focuses on recommendation systems based on social media texts. In particular, the authors propose a Natural Language Processing (NLP) method, called Context-Aware Recommendation Method based on Text and Opinion Mining (CARM-TOM). The method was validated using several experiments with restaurant recommendation data retrieved from the Yelp Internet platform. The proposed CARM-TOM obtained competitive results when compared with a matrix factorization and a state-of-the-art context extraction method. By automatically extracting contextual information, CARM-TOM has the potential to increase the performance of recommendation systems based on opinionated texts.

Areosa and Torgo (2020) discuss a ‘Visual Interpretation of Regression Error’. The paper is set within the hot topic of XAI and it proposes several visualization tools to explain the relationship between input variables and the predictive performance of black box regression ML models. Several illustrations are provided for different real-world datasets and learning algorithms. The proposed visualizations can be used to increase the confidence of decision makers in trusting the regression model predictions.

The sixth and last paper ‘A Bi-objective Procedure to Deliver Actionable Knowledge in Sport Services’, by Pinheiro and Cavique (2020), focuses on a real-world application, aiming to increase the retention of customers in gyms and health clubs. The work presents a data-driven pipeline that is capable of providing actionable knowledge. The authors used data from a Lisbon sports facility that had 21,755 users. A Decision Tree model was then used to predict user dropout rules. Next, several actionable attributes, which can be influenced by the sports facility managers, were identified, aiming to design actionable rules. Finally, several retention intervention what-if scenarios (which use the actionable rules) were simulated, combining ML and causal inference, allowing to estimate their business utility and costs. The overall data-driven actionable approach is thus of potential value to support the decisions of sport facility managers.

Acknowledgments

We would like to thank the other KDBI 2019 track (of EPIA) co-organizers, namely João Gama, Luís Cavique, Manuel Santos and Nuno Marques. Moreover, we wish to thank the authors that contributed with papers for this special issue and the reviewers (from the KDBI 2019 program committee and the EXSY journal). The work of P. Cortez was supported by FCT - *Fundação para a Ciência e Tecnologia* within the R&D Units Project Scope: UIDB/00319/2020.

References

[Ain et al., 2019] Ain, N., Vaia, G., DeLone, W. and Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and suc-

- cess - A systematic literature review. *Decision Support Systems*, 125:113113. doi: 10.1016/j.dss.2019.113113.
- [Andrade et al., 2020] Andrade, T., Cancela, B. and Gama, J. (2020). From Mobility Data to Habits and Common Pathways. *Expert Systems*, pages –.
- [Anjos et al., 2020] Anjos, I., Marques, N. and Grilo, M. (2020). Sibilant consonants classification comparison with multi and single-class neural networks. *Expert Systems*, pages –.
- [Areosa and Torgo, 2020] Areosa, I. and Torgo, L. (2020). Visual Interpretation of Regression Error. *Expert Systems*, pages –.
- [Cortez and Santos, 2013] Cortez, P. and Santos, M. F. (2013). Knowledge Discovery and Business Intelligence. *Expert Systems*, 30(4):283–284. doi: 10.1111/exsy.12042.
- [Cortez and Santos, 2015] Cortez, P. and Santos, M. F. (2015). Recent advances on knowledge discovery and business intelligence. *Expert Systems*, 32(3):433–434. doi: 10.1111/exsy.12087.
- [Cortez and Santos, 2017] Cortez, P. and Santos, M. F. (2017). Third special issue on knowledge discovery and business intelligence. *Expert Systems*, 34(1): e12188. doi: 10.1111/exsy.12188.
- [Cortez and Santos, 2018] Cortez, P. and Santos, M. F. (2018). Fourth special issue on knowledge discovery and business intelligence. *Expert Systems*, 35(4): e12314. doi: 10.1111/exsy.12314.
- [Cortez et al., 2018] Cortez, P., Moro, S., Rita, P., King, D. and Hall, J. (2018). Insights from a text mining survey on Expert Systems research from 2000 to 2016. *Expert Systems*, 35(3): e12280. doi: 10.1111/exsy.12280.
- [Darwiche, 2018] Darwiche, A. (2018). Human-level intelligence or animal-like abilities?. *Communications of ACM*, 61(10):56–67. doi: 10.1145/3271625.
- [Dias et al., 2020] Dias, M., Monteiro, J., Estima, J., Silva, J., Martins, B. (2020). Semantic Segmentation and Colorization of Grayscale Aerial Imagery with W-Net Models. *Expert Systems*, pages –.
- [Fayyad et al., 1996] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [Pinheiro and Cavique, 2020] Pinheiro, P. and Cavique, L. (2020). A Bi-objective Procedure to Deliver Actionable Knowledge in Sport Services. *Expert Systems*, pages –.
- [Sundermann et al., 2020] Sundermann, C., de Padua, R., Tonon, V., Marcacini, M., Domingues, M. and Rezende, S. (2020). A Context-Aware Recommender Method based on Text and Opinion Mining. *Expert Systems*, pages –.

The editors

Paulo Cortez

Paulo Cortez (Habilitation, PhD) is an Associate Professor at the Department of Information Systems, University of Minho, Portugal. He is also assistant director of the ALGORITMI R&D center. His research interests include: Business Analytics (Decision Support, Data Science and Forecasting); and Artificial Intelligence (Machine Learning, Evolutionary Computation and Applications). Currently, he is associate editor of the *Expert Systems* and *Decision Support Systems* journals. He is co-author of more than 170 indexed (e.g., Scopus) publications in international journals (e.g., Decision Support Systems, Information Processing Management) and conferences (e.g., IEEE IJCNN, ACM GECCO). Web-page: <http://www3.dsi.uminho.pt/pcortez>

Albert Bifet

Albert Bifet is Professor at University of Waikato and LTCI, Telecom ParisTech. Previously he worked at Huawei Noah's Ark Lab in Hong Kong, Yahoo Labs in Barcelona, University of Waikato and UPC BarcelonaTech. He is the author of a book on Adaptive Stream Mining and Pattern Learning and Mining from Evolving Data Streams. He is one of the leaders of MOA and Apache SAMOA software environments for implementing algorithms and running experiments for online learning from evolving data streams. He was serving as Co-Chair of the Industrial track of IEEE MDM 2016, ECML PKDD 2015, and as Co-Chair of BigMine (2017, 2016, 2015, 2014, 2013, 2012), and ACM SAC Data Streams Track (2017, 2016, 2015, 2014, 2013, 2012).