Responsible Artificial Intelligence as a Driver of Innovation in Society and Industry*

Alessio Bechini, Alessandro Bondielli, Pietro Ducange, Francesco Marcelloni, Alessandro Renda

Department of Information Engineering
University of Pisa

{alessio.bechini, alessandro.bondielli, pietro.ducange, francesco.marcelloni, alessandro.renda}@unipi.it

Abstract

We describe the most recent research activities of the Artificial Intelligence R&D (AI-RD) group of the Department of Information Engineering of the University of Pisa. The group includes the authors of this contribution (one full professor, two associate professors, and two researchers) along with some PhD students. We also give a glimpse on some of the projects in which the group is involved. Specifically, we focus on those projects where AI is exploited as a driver of innovation in society and industry, especially pointing out some features that led us to label AI as "Responsible".

1 Introduction

The today extensive use of AI calls for responsible AI based on transparency and privacy protection. Transparency is studied in the field of eXplainable AI (XAI) [Arrieta *et al.*, 2020] whose main goal is to generate AI models that are *interpretable*, i.e., that are able to illustrate why the model produces a given output from specific input data (WHAT), and *explainable*, i.e., that are able to explain in a way that is understandable to humans how the model produces a given output from the input data (HOW). The concept of "Explainability" is at the heart of the so-called "Trustworthy AI," which in turn has become central to both users and government agencies: the European Union's "Ethical Guidelines for Trustworthy AI" [EC2, 2019] state that *"[...] AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned"*. Disease diagnosis is a typical example of a setting where explainability plays a key role [Ricatto *et al.*, 2018]: doctors can trust indications from an AI-based system only if they have a chance to check the motivation for the tool outcome, and possibly agree on it.

Many of the algorithms proposed in the literature for learning AI and XAI models are based on the assumption that the training set is statically available in its entirety. Often in practice data are produced in streaming, and thus the learning/updating procedures for the models should be done incrementally over time as the data arrive. This requires appropriate learning algorithms that are able to handle two peculiarities of data streams: i) the sequence of data may be unbounded, thus posing the problem of their storage, and ii) the statistical properties of the data may change over time, due to often unpredictable changes in the underlying non-stationary generation process (a phenomenon identified as "concept drift" in the literature) [Mehmood *et al.*, 2021]. In application scenarios, there are typically multiple data sources used, with each source generating a specific data stream. Obviously, the wider the availability of data, the greater the accuracy and reliability of the learned models. In fact, typically data from different sources reflect different phenomena that might occur over time in contexts where they had not previously been encountered.

Collecting and storing peripheral data from different sources for processing them on a centralized server is often impractical due to two main reasons: on the one hand, it typically introduces severe communication overheads; on the other hand, it violates the privacy and security requirements that are often crucial to data owners. To overcome these limitations, Federated Learning (FL) [Zhang *et al.*, 2021] has been proposed as a privacy preserving paradigm for collaboratively training AI models: in an FL system data owners learn a shared model by only aggregating locally-computed partial models, with no need to share their raw data with other parties.

2 Research Themes

The AI-RD research group has more than 20 years of experience in AI research. In particular, the group has published numerous papers in the field of Fuzzy Rule-based Systems (FRBSs) and Fuzzy Decision Trees (FDTs), with two outstanding papers in two of the most prestigious journals of the field. These models provide *inherent explainability* thanks to the linguistic representation of numerical variables, while still delivering competitive levels of accuracy for both classification and regression tasks. Specifically, we have proposed algorithms to generate FRBSs and FDTs characterized by excellent trade-offs between accuracy and explainability/complexity [Antonelli *et al.*, 2014; Barsacchi *et al.*, 2019]. In the last years, such algorithms

*This work has been partly funded by the European Commission through the H2020 project Hexa-X (Grant Agreement no. 101015956) and by the Italian Ministry of Education and Research (MIUR) in the framework of the CrossLab project (Departments of Excellence).

have also been designed and implemented for being executed on distributed platforms for big data [Barsacchi *et al.*, 2021; Segatori *et al.*, 2018], and ensemble solutions have been investigated as well [Barsacchi *et al.*, 2020]. Recently, the research activity has mainly focused on two mainly topics: FL and machine learning algorithms for streaming data.

2.1 Federated Learning

As an alternative paradigm to data centralization, FL has recently been proposed for collaboratively training an AI model. Importantly, in an FL system data owners do not expose their data to other participants; instead, they learn a shared model by only aggregating locally computed updates. The AI-RD group is proposing solutions for typical data mining problems such as classification/regression and clustering.

We are currently focusing on designing strategies for learning inherently interpretable models (e.g., decision trees, rule based systems) in a federated fashion. Existing approaches for the aggregation of models revolve around the Federated Averaging (FedAvg) protocol [McMahan *et al.*, 2017]; such a protocol is particularly well suited for federated learning of models optimized through Stochastic Gradient Descent, but needs to be revised for other classes of models. In a nutshell, our proposal can be summarized as follows: each data owner builds an XAI model based on local data; then, they share the model with a central server, which is in charge of merging the received models to produce a global model. The model is finally sent back to the data owners that can use it for local inference.

Concerning clustering, we have recently presented a horizontal federated version of the popular k-means and fuzzy c-means algorithms. We have proved that the federated versions are equivalent to the original algorithms applied to the overall dataset [Corcuera Bárcena *et al.*, 2021].

2.2 Machine Learning Algorithms for Streaming Data

In recently published papers we discussed the results coming from the adoption of innovative algorithms for designing descriptive/predictive models for streaming data. The methodological workflow in [Bechini *et al.*, 2020b] leverages an established text classification system for the public opinion assessment from Twitter data. We have considered the case study of the vaccination topic in Italy, and we have reported the results of a Twitter user stance monitoring campaign with advanced analysis on temporal and spatial scales. More recently, we have shown a broad experimental campaign for the same case study, in which we have compared different learning schemes and proposed a novel one, aimed at addressing the event-driven concept drift [Bechini *et al.*, 2021b]. This work stems from the observation that the way people express their opinion and thoughts may change over time according to the event-related landscape that characterizes the Twitter stream in online monitoring applications. We have also recently proposed a novel algorithm for incrementally learning FDTs from streaming data [Ducange *et al.*, 2021]. The algorithm is based on the incremental learning scheme of the well-known Hoeffding Decision Tree, that we modified introducing fuzziness, thus making the tree more robust to noisy and vague data.

As regards descriptive models, in [Bechini *et al.*, 2021a], we have proposed a time-sensitive approach for tracking and profiling changes in the occurrence of notable events in a city, in terms of what happens in the different areas and how possible changes are perceived. We have adopted streams of articles in online newspapers as source of data for monitoring the events in a city and exploited a modified version of the streaming version of the Fuzzy DBSCAN algorithm recently proposed by our group in [Bechini *et al.*, 2020a]. The text of the articles have been processed adopting Sentence-BERT [Reimers e Gurevych, 2019], one of the most recent and powerful techniques for text representation.

3 Funded Projects and Collaborations with Companies

3.1 HEXA-X: The EU Flagship Project for 6G

Even though 5G, the fifth generations of cellular networks, is still in its first stages of commercialization, both academia and industry started to conceive the sixth generation (6G) of cellular networks, which is expected to being deployed starting from 2030. The Hexa-X project, funded by the European Commission within the Horizon 2020 programme[1], provides a flagship vision targeted to lay the foundations of B5G/6G systems [Uusitalo *et al.*, 2021].

AI is expected to play a key role in designing and optimizing the next generation of wireless networks and will be crucial for taming their growing complexity. At the same time, the increasing pervasiveness expected for the future B5G/6G networks in the daily processes of individuals, companies and institutions, makes the need for *trustworthiness* of the native AI components evident. Within the project, our group focuses specifically on addressing the challenges posed by the trustworthiness requirement: in this context, Fed-XAI can be regarded as an enabler for several families of use cases envisioned for 6G, and is proposed as an ML paradigm compliant, at the same time, with *FEDerated learning* and *eXplainable AI* for building trusted and ethical networks. A related demonstration activity will aim to develop and validate a framework for Fed-XAI: practically, end-user agents will sense local data and collaboratively train an XAI model with the goal to make predictions on the Quality of Experience (QoE) provided by the network, without exposing their data to other agents. Recently, in a preliminary comparative analysis, we have investigated the trade-off between accuracy and explainability in the context of tree-based models (e.g., fuzzy decision trees and random forests) for QoE prediction [Renda *et al.*, 2021].

3.2 The AUTOMIA project

The AUTOMIA project focuses on the development of a *Robotic Process Automation* (RPA) system for small and medium-sized companies. An RPA aims to automatize low-level tasks that are repetitive and rather prone to potential errors. In the project a key aspect is a *responsible use of AI*

[1]Horizon 2020 - G.A.: 101015956

and data mining tools, to benefit workers (rather than replacing them), by supporting them with a tool to automatically recognize recurring patterns in tasks and sub-tasks, and enable their automation. Thus, the workers could tackle more challenging and rewarding tasks and avoid errors. To achieve these objectives, we aim to exploit data mining algorithms for Sequential Pattern Mining and Frequent Episode Mining [Mannila *et al.*, 1997], as well as to employ machine learning to optimize the sequence of automated operations.

3.3 The Experience of Co-Designing a Machine Learning Software for Document Categorization

In the framework of a collaboration with an Italian company, leader in field of processing and storage of digital documents, the AI-RD group has co-designed a software module that exploits innovative solutions in text mining and machine learning. The module has been integrated into the company's IT infrastructure, to automatically identify the specific type of a request or complaint, coming from a customer through a paper letter, and select the most suitable group of operators for its handling. In the first place, the researchers of the AI-RD group delivered a technical course to a small group of employees of the company (programmers and software analysts) focusing on Machine Learning and Text Mining. At the same time, a deep analysis of solutions previously adopted by the company has been performed, deriving new requirements for the text categorization module. Then, upon a thorough design exploration, the most appropriate model was actually embedded in the software module. Finally, after a testing period, it was finally deployed in the IT infrastructure and used in production mode. An analysis of the company's KPIs has shown a 50% increase in automation, compared to the previous tools used for the management of requests.

3.4 IT2PAO: A Research and Development Lab on Intelligent Techniques for Process Automation and Optimization

The AI-RD group is currently experimenting a new type of collaboration with LogObject AG[2], a Swiss company, through the activation of a joint research and development laboratory. The lab is managed by a joint scientific board, which annually approves a plan of activities and the necessary human resources (PhD students, PostDocs and young researchers) funded by the company for performing them. The activities of the laboratory focus on applications of AI for process automation and optimization, with particular attention to human factors, as the target processes typically involve scheduling, monitoring, and coordination of tasks performed by specialized personnel. Clearly, this application field requires the adoption of "responsible" AI.

4 Software

4.1 SK-MOEFS

In [Gallo *et al.*, 2020] we have described the software architecture of a Python tool for generating a set of FRBSs characterized by different trade-off between accuracy and explainability. The tool has been implemented so as to be fully integrated into Scikit-Learn library, the most famous and used Machine Learning Python library. The tool offers the implementation of PAES-RCS [Antonelli *et al.*, 2014], a multi-objective evolutionary learning scheme for generating explainable and accurate fuzzy rule-based classifiers. Moreover, the tool can be easily extended for integrating new types of fuzzy models and different learning schemes. The library is publicly available on a GitHub repository[3].

4.2 Fuzzy Decision Trees

The AI-RD group has publicly released two different versions of a tool for generating FDTs for classification tasks. The first version[4] contains distributed algorithms, implemented in SCALA, for generating both binary and multi-nominal FDTs from big datasets. The second version [5] is a sequential implementation in Python for quickly experimenting FDTs with datasets with a moderate dimension, without the necessity of setting up a cluster of servers. Both versions also include a fuzzy discretizer for automatically generating a fuzzy partition from data for each input variable.

4.3 Native FED-XAI Application for 6G Networks

This application is currently under development as part of the Hexa-X project. It will allow connected devices to exploit the Fed-XAI functionality, possibly exploiting the Multi-access Edge Computing (MEC) architecture for cloud computing capability at the edge of the network [Noferi *et al.*, 2021]. It will feature three macro-modules: a *local user application* module, a Fed-XAI module for client-side and server-side components of the FL process, and a FED-XAI Dashboard module to let any concerned stakeholder check explanations of predictions.

5 Challenges and Perspectives

Federated approaches to learn XAI models pose several challenges, especially concerning the "responsible AI" facet of systems designed to leverage relative frameworks. First of all, explainability preservation in aggregation of local XAI models has not been satisfactorily explored and needs a deeper analysis. Second, nodes often collect streaming data. This requires a continuous update of the local models and consequently of the aggregated model. Third, concerning clustering, horizontal federated learning for some classical clustering algorithms has already been proposed. On the other hand, vertical federated clustering has not been sufficiently explored yet.

References

[Antonelli *et al.*, 2014] Michela Antonelli, Pietro Ducange, e Francesco Marcelloni. A fast and efficient multi-objective evolutionary learning scheme for fuzzy rule-based classifiers. *Information Sciences*, 283:36–54, 2014.

[2]https://logobject.com

[3]https://github.com/GionatanG/skmoefs

[4]https://github.com/BigDataMiningUnipi/FuzzyDecisionTreeSpark

[5]https://bitbucket.org/mbarsacchi/fuzzyml/src/master/

[Arrieta *et al.*, 2020] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion*, 58:82–115, 2020.

[Barsacchi *et al.*, 2019] Marco Barsacchi, Alessio Bechini, Pietro Ducange, e Francesco Marcelloni. Optimizing partition granularity, membership function parameters, and rule bases of fuzzy classifiers for big data by a multi-objective evolutionary approach. *Cognitive Computation*, 11(3):367–387, 2019.

[Barsacchi *et al.*, 2020] Marco Barsacchi, Alessio Bechini, e Francesco Marcelloni. An analysis of boosted ensembles of binary fuzzy decision trees. *Expert Systems with Applications*, 154:113436, 2020.

[Barsacchi *et al.*, 2021] Marco Barsacchi, Alessio Bechini, e Francesco Marcelloni. Implicitly distributed fuzzy random forests. In *Proc. of the 36th Annual ACM Symposium on Applied Computing*, page 392–399, New York, NY, USA, 2021. Association for Computing Machinery.

[Bechini *et al.*, 2020a] A. Bechini, F. Marcelloni, e A. Renda. TSF-DBSCAN: a Novel Fuzzy Density-based Approach for Clustering Unbounded Data Streams. *IEEE Transactions on Fuzzy Systems*, pages 1–1, 2020.

[Bechini *et al.*, 2020b] Alessio Bechini, Pietro Ducange, Francesco Marcelloni, e Alessandro Renda. Stance analysis of twitter users: the case of the vaccination topic in italy. *IEEE Intelligent Systems*, 2020.

[Bechini *et al.*, 2021a] Alessio Bechini, Alessandro Bondielli, José Luis Corcuera Bárcena, Pietro Ducange, Francesco Marcelloni, e Alessandro Renda. Mining the stream of news for city areas profiling: a case study for the city of rome. In *2021 IEEE Int'l Conf. on Smart Computing (SMARTCOMP)*, pages 317–322. IEEE, 2021.

[Bechini *et al.*, 2021b] Alessio Bechini, Alessandro Bondielli, Pietro Ducange, Francesco Marcelloni, e Alessandro Renda. Addressing event-driven concept drift in twitter stream: a stance detection application. *IEEE Access*, 2021.

[Corcuera Bárcena *et al.*, 2021] José Luis Corcuera Bárcena, Francesco Marcelloni, Alessandro Renda, Alessio Bechini, e Pietro Ducange. A federated fuzzy c-means clustering algorithm. In *Proc. of the 13th Int'l Workshop on Fuzzy Logic and Applications (WILF 2021)*, 2021.

[Ducange *et al.*, 2021] Pietro Ducange, Francesco Marcelloni, e Riccardo Pecori. Fuzzy hoeffding decision tree for data stream classification. *Int'l Journal of Computational Intelligence Systems*, 14(1):946–964, 2021.

[EC2, 2019] Ethics Guidelines for Trustworthy AI, Technical Report, 2019. European Commission. High Level Expert Group on AI. https://bit.ly/3AdnloI.

[Gallo *et al.*, 2020] Gionatan Gallo, Vincenzo Ferrari, Francesco Marcelloni, e Pietro Ducange. SK-MOEFS: A library in python for designing accurate and explainable fuzzy models. In *Int'l Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 68–81. Springer, 2020.

[Mannila *et al.*, 1997] Heikki Mannila, Hannu Toivonen, e A Inkeri Verkamo. Discovery of frequent episodes in event sequences. *Data mining and knowledge discovery*, 1(3):259–289, 1997.

[McMahan *et al.*, 2017] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, e Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282. PMLR, 2017.

[Mehmood *et al.*, 2021] Hassan Mehmood, Panos Kostakos, Marta Cortes, Theodoros Anagnostopoulos, Susanna Pirttikangas, e Ekaterina Gilman. Concept drift adaptation techniques in distributed environment for real-world data streams. *Smart Cities*, 4(1):349–371, 2021.

[Noferi *et al.*, 2021] Alessandro Noferi, Giovanni Nardini, Giovanni Stea, e Antonio Virdis. Deployment and configuration of mec apps with simu5g. *arXiv preprint arXiv:2109.12048*, 2021.

[Reimers e Gurevych, 2019] Nils Reimers e Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019.

[Renda *et al.*, 2021] Alessandro Renda, Pietro Ducange, Gionatan Gallo, e Francesco Marcelloni. Xai models for quality of experience prediction in wireless networks. In *2021 IEEE Int.l Conf. on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2021.

[Ricatto *et al.*, 2018] Mattia Ricatto, Marco Barsacchi, e A. Bechini. Interpretable cnv-based tumour classification using fuzzy rule based classifiers. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, page 54–59, New York, NY, USA, 2018. Association for Computing Machinery.

[Segatori *et al.*, 2018] Armando Segatori, Alessio Bechini, Pietro Ducange, e Francesco Marcelloni. A distributed fuzzy associative classifier for big data. *IEEE Transactions on Cybernetics*, 48(9):2656–2669, 2018.

[Uusitalo *et al.*, 2021] Mikko A Uusitalo, Mårten Ericson, Björn Richerzhagen, Elif Ustundag Soykan, Patrik Rugeland, Gerhard Fettweis, Dario Sabella, Gustav Wikström, Mauro Boldi, Marie-Helene Hamon, et al. Hexa-x the european 6g flagship project. In *2021 Joint European Conf. on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, pages 580–585. IEEE, 2021.

[Zhang *et al.*, 2021] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, e Yuan Gao. A survey on federated learning. *Knowledge-Based Systems*, 216:106775, 2021.