Federated Learning of XAI Models in 6G Systems

Subjects: Others

Contributor: Alessandro Renda, Pietro Ducange, Francesco Marcelloni, Dario Sabella, Miltiadis C. Filippou, Giovanni Nardini, Giovanni Stea, Antonio Virdis, Davide Micheli, Damiano Rapone, Leonardo Gomes Baltar

The federated learning (FL) of neural networks has been widely investigated exploiting variants of stochastic gradient descent as the optimization method, it has not yet been adequately studied in the context of inherently explainable models. On the one side, eXplainable Artificial Intelligence (XAI) permits improving user experience of the offered communication services by helping end users trust (by design) that in-network AI functionality issues appropriate action recommendations. On the other side, FL ensures security and privacy of both vehicular and user data across the whole system.

Keywords: explainable artificial intelligence ; federated learning ; 6G

1. Introduction

Artificial Intelligence (AI), along with Machine Learning (ML) as one of its core building blocks, is entering many market domains at a fast pace and will not only leverage advanced communication networks but also shape the definition of next-generation networks themselves. In particular, AI is expected to play a crucial role in the design, operation and management of future beyond-5G (B5G)/6G networks and in a plethora of applications [1]. However, the introduction of in-network AI comes with growing concerns on privacy, security and trust for citizens and users; for this reason, the adoption of eXplainable AI (XAI) models is an emerging trend considered for the design of transparent AI-based solutions. Moreover, future service scenarios, especially in the automotive domain, will be characterized by the deployment of connected vehicular systems from heterogeneous car manufacturers, connected via different Mobile Network Operators (MNOs) and different technology infrastructures [2]. In such complex setups, it will be imperative for service providers to consider federated network environments including multiple administrative and technical domains as a working assumption for the design of innovative applications. It is worth noting that the automated driving use case of “Teleoperated Driving (ToD) for Remote Steering” [3] requires a throughput of up to 36 Mbps per single stream, along with a positioning accuracy of 0.1 m and a reliability of 99.999% for the service to be considered available to the end customer. Such stringent requirements call for new technical enablers, to be introduced as part of the 6G network design. Considering the above-mentioned challenges, researchers envision the use of the federated learning (FL) concept applied jointly with XAI models and discuss its applicability to automated vehicle networking use cases to be encountered in B5G/6G setups. In fact, although FL has recently been widely investigated in the context of Neural Networks and Deep Learning models (due to their gradient based optimization strategy), much less attention has been devoted so far to FL of XAI models.

1.1. The Need for XAI

The adoption of AI techniques cannot disregard the fundamental value of trustworthiness, which, along with inclusiveness and sustainability, represents the three core values of the European Union Flagship Hexa-X (www.hexa-x.eu (accessed on 16 August 2022)) vision for the upcoming 6G era [1]. Trustworthiness has become paramount for both users and government entities, as witnessed by the “right to explanation” described in the General Data Protection Regulation (GDPR) and by the European Commission’s (EC) Technical Report on “Ethics guidelines for trustworthy AI” [4]. According to these, explainability represents a key requirement towards trustworthiness. Thus, industry and academia are placing increasing attention on XAI, that is, an AI “that produces details or reasons to make its functioning clear or easy to understand” [5].

In this context, two strategies for achieving explainability can be identified [5]: the adoption of post-hoc explainability techniques (i.e., the “explaining black-box” strategy) and the design of inherently interpretable models (i.e., “transparent box design” strategy). Researchers focus on this latter class of approaches, noting that certain applications may tolerate a limited performance degradation to achieve fully trustworthy operation. In fact, performance and transparency are typically considered conflicting objectives [5][6]. However, this trade-off holds as long as the target task entails a certain complexity and the data available are many and high quality. In this case, complex models, such as Deep Neural Networks (DNNs), which are hard to interpret due to their huge number of parameters and non-linear modelling, have proven to achieve high levels of accuracy; conversely, decision trees and rule-based models may feature lower modelling capability but are typically considered “highly interpretable”.

The importance of explainability has been recently highlighted in the context of Secure Smart Vehicles [7]: on one hand, explanation is crucial in safety-critical AI-based algorithms, designed to extend some widely available capabilities (e.g., lane-keeping and braking assistants) towards fully automated driving; on the other hand, explainability is needed at the design stage to perform model debugging and knowledge discovery, thus positively impacting system security by reducing model vulnerabilities against external attacks. Explainability of AI models will be crucial for 6G-enabled V2X systems. A prime example is an AI service consumer requesting in-advance notifications on QoS predictions, as studied in Hexa-X [1] and the 5G Automotive Association (5GAA) [8]. Accurate and timely predictions should support very demanding use cases, with a horizon ranging from extremely short to longer time windows. Better explainability of such predictions and any consequent decision will provide benefits not only for technology and service providers, but also for end-customers, who will become more receptive to AI-based solutions.

1.2. Federated Learning

Exploiting data from multiple sources can enhance the performance (i.e., high accuracy based on reduced bias) of AI models. However, wirelessly collecting and storing peripheral data for processing on a centralized server has become increasingly impractical due to two main reasons: first, it typically introduces severe communication and computation overhead due to the transmission and storage of large training data sets, respectively; second, it violates the privacy and security requirements imposed by data owners by expanding the surface of possible over-the-air attacks towards biased decision making. In other words, the preservation of data privacy represents an urgent requirement of today’s AI/ML systems, because data owners are often reluctant to share their data with other parties; in some jurisdictions, users have the ability to consent or not with the sharing of privacy-sensitive data (e.g., per the General Data Protection Regulation—GDPR in European Union). Such a need to preserve privacy of data owners, however, clashes with the need to collect data to train accurate ML models, which are typically data hungry in their learning stage. To overcome these limitations, FL has been proposed as a privacy-preserving paradigm for collaboratively training AI models. In an FL system, participants iteratively learn a shared model by only transferring local model updates and receiving an aggregated shared model update, without sharing raw data.

The main opportunities of FL in the context of Intelligent Transportation Systems (ITS) have been recently discussed in [9]: FL is expected to support both vehicle management (i.e., automated driving) and traffic management (i.e., infotainment and route planning) applications. Furthermore, FL has been applied in the context of Ultra-Reliable Low-Latency Communications for Vehicle-to-Vehicle scenarios, allowing vehicular users to estimate the distribution of extreme events (i.e., network-wide packet queue lengths exceeding a predefined threshold) with a model learned in a decentralized manner [10]. The model parameters are obtained by executing maximum likelihood estimation in a federated fashion, without sharing the local queue state information data. The concept of Federated Vehicular Network (FVN) has been recently introduced [11], as an architecture with decentralized components that natively support applications, such as entertainment at sport venues and distributed ML. However, FVN is a stationary vehicular network and relies on the assumption that vehicles remain at a fixed location, e.g., parking lots, so that the wireless connection is stable.

2. FED-XAI: Bringing Together Federated Learning and Explainable AI

Existing AI-based solutions for wireless network planning, design and operation ignore either or both of the following aspects: (i) the need to preserve data privacy at all times, including wireless transfer and storage, and (ii) the explainability of the involved models. Furthermore, latency and reliability requirements of safety-critical automotive communications call for seamless availability of decentralized and lightweight intelligence, where data are generated—and decisions made—anytime and anywhere.

Current FL approaches only address the first requirement. Explainability has been given less attention, having been approached primarily by exploiting post-hoc techniques, e.g., Shapley values to measure feature importance [12]. There is a substantial lack of approaches for FL of inherently explainable models. On the other hand, a federated approach for learning interpretable-by-design models, in which transparency is guaranteed for every decision made, would represent a significant leap towards trustworthy AI. Therefore, researchers introduce the concept of FL of XAI (FED-XAI) models, as a framework with a twofold objective: first, to leverage FL for privacy preservation during collaborative training of AI models, especially suitable in heterogeneous B5G/6G scenarios; second, to ensure an adequate degree of explainability of the models themselves (including the obtained aggregated model as a result of FL).

First, it is worth noting that standard algorithms for learning such models typically adopt a heuristic approach; in fact, gradient descent-based optimization methods, widely used in FL, cannot be immediately applied, as they require the formulation of a global objective function. The greedy induction of decision trees, for example, recursively partitions the feature space by selecting for each decision node the most suitable attribute. The major challenge of the FED-XAI approach, therefore, consists in generating XAI models, whose FL is not based on the optimization of a differentiable global objective function.

The proposed FED-XAI approach relies on orchestration by a central entity but ensures that local data are not exposed beyond source devices: each data owner learns a model by elaborating locally acquired raw data and shares such a model with the central server, which merges the received models to produce a global model (**Figure 1**). Notably, the envisioned approach for federated learning of explainable AI models ensures data privacy regardless of the data sample size. As per the advantages of the FED-XAI approach, researchers expect that the global aggregated model performs better than the local models because it exploits the overall information stored and managed by all data owners, without compromising model interpretability.

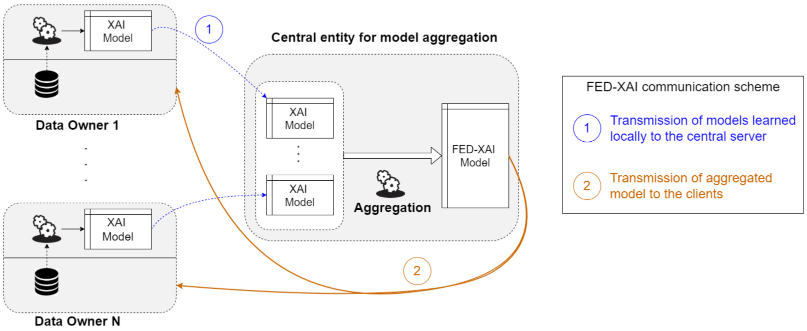
XAI Model
1
Central entity for model aggregation
Data Owner 1
XAI Model
XAI Model
FED-XAI Model
Aggregation
XAI Model
Data Owner N
2
FED-XAI communication scheme
1 Transmission of models learned locally to the central server
2 Transmission of aggregated model to the clients

Figure 1. Illustration of federated learning of XAI models.

As a consequence, the communication overhead is reduced, and the system is more robust to possible connectivity problems. Second, merging decision trees and rule-based models requires defining appropriate procedures, necessarily different from the simple weighted average of models of the FedAvg protocol applied, for example, to NNs. In more detail, the XAI models researchers consider can be represented as collections of “IF *antecedent* THEN *consequent*” rules, (natively in a rule-based system, and easily obtainable also from a decision tree). This representation is applicable regardless of the target task (regression or classification) and the type of the attributes (e.g., nominal or numeric). The aggregation procedure consists in juxtaposing rules collected from data owners, and resolving possible *conflicts*, which emerge when rules from different models, having antecedents referring to identical or overlapping regions of the attribute space, have different consequents. In one of the recent works [13], researchers presented a novel approach for FL of Takagi-Sugeno-Kang (TSK) fuzzy rule based systems [14], which can be considered as XAI models in regression problems. In a TSK model, the antecedent of a rule identifies a specific region of the attribute space, whereas the corresponding consequent allows for the evaluation of the predicted output within such a region as a linear combination of the input variables. When two rules, generated by different clients, share the same antecedent, the aggregation strategy for generating the FED-XAI model involves combining the two rules into a single one with the same antecedent: the coefficients of the linear model of the new consequent are evaluated as the weighted average of the coefficients of the original rules, where the weight of each rule depends on its support and confidence values. Research efforts in the FED-XAI domain, however, are still in their embryonic stage: as for tree-based models, a preliminary investigation of the trade-off between accuracy and interpretability has been recently carried out [15], but learning strategies compliant with the federated setting still need to be sharpened.

Main Challenges of the FED-XAI Approach

There are also challenges related to the FED-XAI approach, especially for time-critical operations in automated driving setups. For example, the computation (and, therefore, energy) footprint of FED-XAI needs to be pre-evaluated before implementation to identify the scalability potential of the solution. A clear distinction should be made between the stages of *training* and *inference*. For most ML models, including decision tree and rule-based systems, the inference time (critical from automated driving service standpoint) is negligible compared to the training time and, in any case, model complexity can be tuned to ensure that time constraints are satisfied. A larger computational overhead is required in the training stage, but it does not affect the application (e.g., learning can be performed in idle state). Another challenge is FED-XAI system resilience to attackers trying to benefit from the access to explanations of QoE predictions (e.g., towards increasing automated driving service outages for all or targeted vehicles). Finally, the approach will also need to address some additional challenges that are typical of FL and are likely to characterize 6G network-based intelligent transportation applications: (i) multi-source data may have different distributions and volumes, (ii) the number of participants can grow fast and their participation to FL may be unstable due to insufficiency of radio and computational resources, and (iii) learned models will need to be agilely updated in scenarios where concept drift alters the characteristics of data distributions over time.

References

1. Hexa-X Deliverable D1.2—Expanded 6G Vision, Use Cases and Societal Values—Including Aspects of Sustainability, Security and Spectrum. Available online: https://hexa-x.eu/d1-2-expanded-6g-vision-use-cases-and-societal-values-including-aspects-of-sustainability-security-and-spectrum/ (accessed on 3 May 2021).
2. 5GAA Working Item MEC4AUTO. Technical Report Use Cases and Initial Test Specifications Review. Available online: https://5gaa.org/news/working-item-mec4auto/ (accessed on 19 July 2021).
3. 5GAA Technical Report. Tele-Operated Driving (ToD): System Requirements Analysis and Architecture. Available online: https://5gaa.org/news/tele-operated-driving-tod-system-requirements-analysis-and-architecture/ (accessed on 15 September 2021).
4. Ethics Guidelines for Trustworthy AI, Technical Report. European Commission. High Level Expert Group on AI. 2019. Available online: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (accessed on 16 August 2022).
5. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 2020, 58, 82–115.
6. Fernandez, A.; Herrera, F.; Cordon, O.; Jose del Jesus, M.; Marcelloni, F. Evolutionary Fuzzy Systems for Explainable Artificial Intelligence: Why, When, What for, and Where to? IEEE Comput. Intell. Mag. 2019, 14, 69–81.
7. Scalas, M.; Giacinto, G. On the Role of Explainable Machine Learning for Secure Smart Vehicles. In Proceedings of the 2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), Turin, Italy, 18–20 November 2020; pp. 1–6.
8. 5GAA White Paper: Making 5G Proactive and Predictive for the Automotive Industry. White Paper. Available online: https://5gaa.org/news/5gaa-releases-white-paper-on-making-5g-proactive-and-predictive-for-the-automotive-industry/ (accessed on 8 January 2020).
9. Elbir, A.M.; Soner, B.; Coleri, S. Federated learning in vehicular networks. arXiv 2020, arXiv:2006.01412.
10. Samarakoon, S.; Bennis, M.; Saad, W.; Debbah, M. Federated Learning for Ultra-Reliable Low-Latency V2V Communications. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 9–13 December 2018; pp. 1–7.
11. Posner, J.; Tseng, L.; Aloqaily, M.; Jararweh, Y. Federated Learning in Vehicular Networks: Opportunities and Solutions. IEEE Netw. 2021, 35, 152–159.
12. Salim, S.; Turnbull, B.; Moustafa, N. A Blockchain-Enabled Explainable Federated Learning for Securing Internet-of-Things-Based Social Media 3.0 Networks. IEEE Trans. Comput. Soc. Syst. 2021, 1–17.
13. Corcuera Bárcena, J.L.; Ducange, P.; Ercolani, A.; Marcelloni, F.; Renda, A. An Approach to Federated Learning of Explainable Fuzzy Regression Models. In Proceedings of the IEEE WCCI 2022 (World Congress on Computational Intelligence), Padua, Italy, 18–23 July 2022.
14. Takagi, T.; Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Syst. Man Cybern. 1985, SMC-15, 116–132.
15. Bechini, A.; Corcuera Bárcena, J.L.; Ducange, P.; Marcelloni, F.; Renda, A. Increasing Accuracy and Explainability in Fuzzy Regression Trees: An Experimental Analysis. In Proceedings of the IEEE WCCI 2022 (World Congress on Computational Intelligence), Padua, Italy, 18–23 July 2022.