EFLChain: Ensemble Learning via Federated Learning over Blockchain network: a framework

Umer Majeed and Choong Seon Hong

Department of Computer Engineering, Kyung Hee University, Yongin, 446-701 Korea
{umermajeed, cshong }@khu.ac.kr

Abstract

Ensemble learning methods have superior predictive performance, enhanced robustness, and high accuracy but reduced variance comparative to stand-alone learning models. Federated learning is a distributed method for collaborative training of learning models without the need for training data to be centralized. However, federated learning with blockchain further improvises the security of the learning process by rectifying the central server as a single point of failure. In this paper, we propose a framework for ensemble learning where the baseline global models are trained on the blockchain via federated learning (FLchain). The baseline global models are ensembled in the ensemble layer to devise the final ensemble model.

1. Introduction

Ensemble learning is the methodology to contrive a strong model by combining several weaker baseline models [1]. The predictive performance of ensemble models is found to be higher and robust than constituent baseline models alone [2]. The baseline models are usually simultaneously trained on the same training set. With different underlying machine-learning algorithms, diversity in values of hyper-parameters, improved accuracy can be achieved for the same training and testing sets. The base models may be neural networks with a different number of hidden layers, decision trees, support vector machine (SVM) and k-nearest neighbor. Each of the algorithms may learn a diverse aspect of training data. Ensemble methods such as data partition-based ensembles, model-based ensembles have found to be more precise and robust relative to single models.

Federated learning [3][4] is the distributed machine learning approach where the privacy of user data remained preserved. User devices train local model updates on their data. These local model updates are aggregated on a central server to build the global model. This approach towards machine learning does not require raw data to be accumulated at a central server. Federated learning is efficient with regard to privacy, latency, energy, storage and performance perspective. Federated learning has been shown to remain stable under non-i.i.d. (not independent and identically distributed) training data.

Researchers are integrating blockchain with federated learning to mitigate the single point of failure at the central server. Instead, the blockchain holds the global model in an immutable manner [5]. Moreover, the provenance of local model updates is maintained in the distributed ledger. The permissioned blockchain assisted federated learning alleviates the poisoning attacks [6][7] from malicious agents. A blockchain-enabled federated learning scheme called “FLchain” is proposed in [8].

In this paper, we propose the framework for ensemble learning via federated learning over the blockchain network. User devices offer the same training data for training of multiple global baseline models via federated learning for subsequent ensemble learning. We use FLchain for the training of baseline global models.

2. Proposed Framework

In this section, we propose “EFLChain”: a framework for ensemble learning, where the baseline models are trained using federated learning over the blockchain

network. For the training of the baseline models, we employ FLchain [8]. The layered architecture of the proposed framework is shown in Fig. 1. FLchain can be used for training of base models which can be later ensembled for higher accuracy.

The working of the "EFLChain" framework is explained below:

- **Initialization**: The global baseline models are aggregated on the blockchain layer. Specifically, for each baseline model in ensemble learning, a new channel is assigned. Each channel will provide a global base model after the federated learning process via FLchain. The initial configuration for each baseline model is stored in the genesis block of the corresponding channel.
- **Device Layer**: The device layer comprises several user's mobile devices that agree to take part in the ensemble learning paradigm for a particular learning problem. These devices are carefully selected based on criteria defined by data quality, availability, battery power and computation capability [9][10]. Each device has training data which offers significant utility for the learning problem. Each device is assigned a pair of private-public keys. The private key is used to digitally sign the local model updates to guarantee provenance.
  - **Local Layer**: For a user's mobile device, the local layer contains all local baseline models which are trained on the same data. The local models are trained.
- **Edge layer:** The local model updates are sent by the user's mobile devices in the device layer to edge servers in the edge layer.

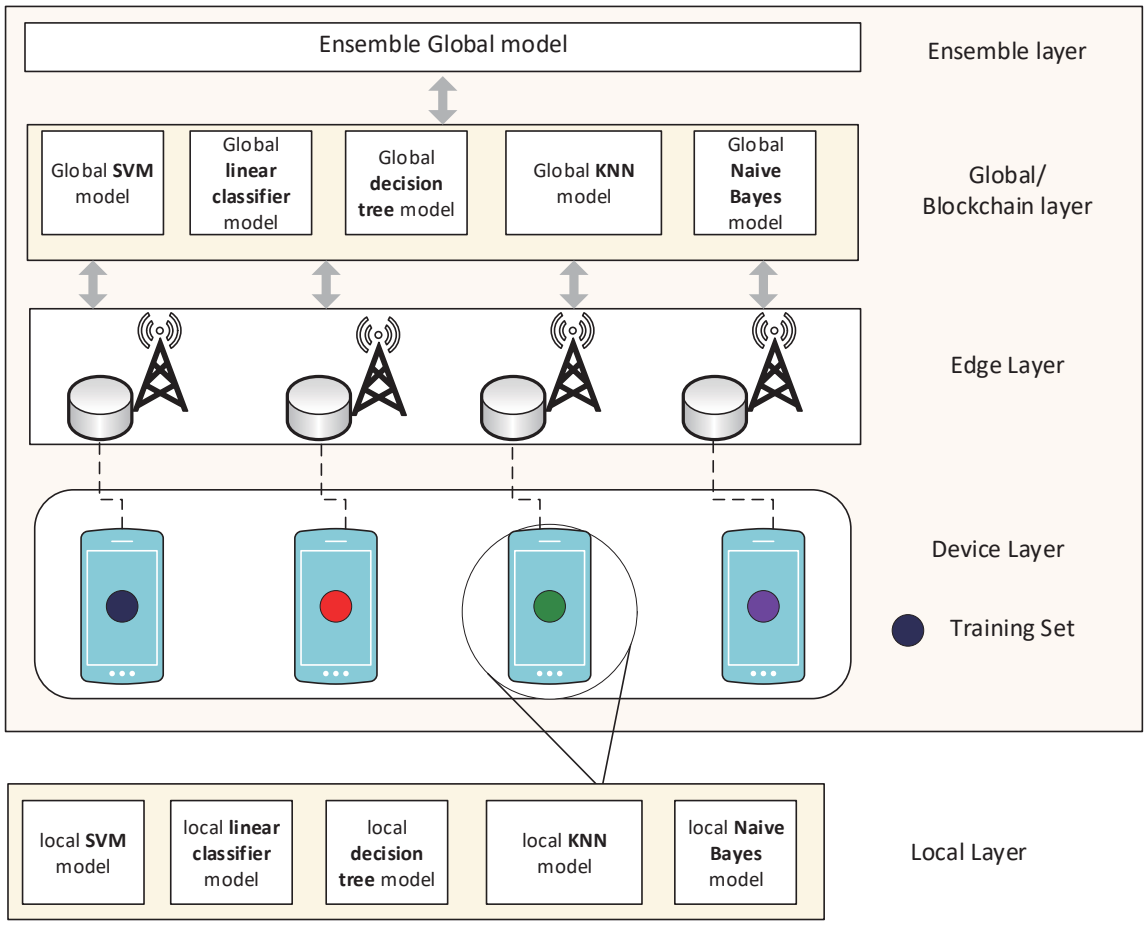

Figure 1: EFLChain : Framework for ensemble learning via federated learning over Blockchain Network

- **Blockchain/Global Layer:** Blockhain layer constitutes of multiple global models as required by ensembling learning problem**.** The local model updates are aggregated on the blockchain layer for each global baseline model. The aggregation procedure is described in FLchain [8] for a global model. Specifically, for each global iteration for a global model, the local updates are bundled in the block, the global model state trie is updated [8] and the newly formed block is appended on channel-specific layer. After a pre-defined stop condition is achieved, the global model state trie has the weight parameters of the global baseline model.
- **Ensemble Layer**: Global base models can be ensembled off-chain for higher precision of the resultant learning model at the ensemble layer. The ensemble algorithm is specified as required by learning problem i.e. simple or weighted averaging for regression, and simple or weighted majority-voting for classification.

3. Conclusion

Ensemble learning techniques have been demonstrated to show improved performance relative to single models. In this study, we presented the framework "EFLchain" for ensemble learning via blockchain assisted federated learning. Global base models are trained seamlessly on the same data of user devices. Each global base model is considered independent and possesses its own channel and channel-specific ledger on the blockchain layer. Subsequently, global baseline models are ensembled off-chain to devise the ensemble model.

Acknowledgment:

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-01287, Evolvable Deep Learning Model Generation Platform for Edge Computing), and by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2015-0-00567, Development of Access Technology Agnostic Next-Generation Networking Technology for Wired-Wireless Converged Networks) and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-01287, Evolvable Deep Learning Model Generation Platform for Edge Computing). *Dr. CS Hong is the corresponding author.

[REFERENCES]

[1] Dietterich, Thomas G. "Ensemble methods in machine learning." *International workshop on multiple classifier systems*. Springer, Berlin, Heidelberg, 2000.

[2] Zheng, Jiewan, et al. "Deep ensemble machine for video classification." *IEEE transactions on neural networks and learning systems* 30.2 (2018): 553-565.

[3] H. B. McMahan, and D. Ramage, "Federated Learning: Collaborative Machine Learning without Centralized Training Data," [Online] Available at: https://ai.googleblog.com/2017/04/federated-learningcollaborative.html, 2017.

[4] Tran, Nguyen H., et al. "Federated Learning over Wireless Networks: Optimization Model Design and Analysis." *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019.

[5] Kim, Hyesung, et al. "Blockchained on-device federated learning." *IEEE Communications Letters* (2019).

[6] B. Biggio, B. Nelson, and P. Laskov. Poisoning attacks against support vector machines. In ICML, 2012.

[7] Zhao, Yang, et al. "Mobile Edge Computing, Blockchain and Reputation-based Crowdsourcing IoT Federated Learning: A Secure, Decentralized and Privacy-preserving System." *arXiv preprint arXiv:1906.10893* (2019).

[8] Majeed, Umer, and Choong Seon Hong. "FLchain: Federated Learning via MEC-enabled Blockchain Network." in 2019 *20th IEEE Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pp. 1-4, 2019.

[9] Lim, Wei Yang Bryan, et al. "Federated Learning in Mobile Edge Networks: A Comprehensive Survey." *arXiv preprint arXiv:1909.11875* (2019).

[10] Khan, Latif U., et al. "Federated Learning for Edge Networks: Resource Optimization and Incentive Mechanism." *arXiv preprint arXiv: 1911.05642* (2019).