Adaptive Aggregation Weight Assignment for Federated Learning: A Deep Reinforcement Learning Approach

Extended Abstract

Enwei Guo
School of Computer Science and Engineering, South China University of Technology
Guangzhou, China
cs_gew@mail.scut.edu.cn

Xiumin Wang
School of Computer Science and Engineering, South China University of Technology
Guangzhou, China
xmwang@scut.edu.cn

Weiwei Wu
School of Computer Science and Engineering, Southeast University
Nanjing, China
weiweiwu@seu.edu.cn

ABSTRACT

Federated learning (FL) has recently received considerable attention due to its capability of allowing distributed clients to collaboratively train a global model without sharing their private data. However, due to the heterogeneous data distributions/contents of clients, it is non-trivial to accurately evaluate the contributions of local models in global model aggregation. Most existing works simply use the amount of data to assign weights of clients in global model aggregation, e.g., FedAvg, which however is not effective, especially when the data of clients is non-IID. To address this issue, this paper aims to propose a novel FL algorithm, which can accurately evaluate the contributions of clients and aggregate global model in a more efficient manner. More specifically, this paper proposes a Deep Reinforcement Learning (DRL) method to dynamically learn the contributions of clients in each communication round. Based on this, we adaptively assign appropriate weights to clients, which will be used in global model aggregation. By improving the process of global mode aggregation, our proposed scheme greatly improves the performance of federated learning.

KEYWORDS

Federated Learning; Deep Reinforcement Learning; Weight Assignment

ACM Reference Format:

Enwei Guo, Xiumin Wang, and Weiwei Wu. 2022. Adaptive Aggregation Weight Assignment for Federated Learning: A Deep Reinforcement Learning Approach: Extended Abstract. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 INTRODUCTION

With the advent of Internet of Things (IoT) [1], a large volume of data has been recently generated at local intelligent devices. These big data can be utilized to train machine learning models, and obtain useful information, e.g., the detection, classification and the prediction of future events. However, due to the limited network bandwidth and data privacy concerns, it is impractical to collect all the users' data to a central server for further processing.

To address the above challenges, *Federated Learning*, also known as Federated Optimization, has recently been proposed by Google [2–4]. Roughly speaking, FL allows multiple parties to collaboratively train a machine learning model without data sharing and collection, which particularly fits well in IoT. FL has recently received considerable attention in IoT and edge computing [5]. For instance, [9] uses FL to improve the next word predictions of Google keyboard. In [10, 11], FL is applied to the medical field.

Despite the great potential, FL is still facing several challenges. One of the most critical issues is that the data residing on clients might not be representative of the population distribution, i.e., non-independent and identically distribution (non-IID), due to the heterogenous interests of users' devices. Such non-IID data distribution will inevitably bring the biases in model training, and cause accuracy degradation. To alleviate the accuracy loss incurred by non-IID data, several research efforts have been devoted in the literature[6–8]. Of course, accurately assessing the contribution of local models to global model can also greatly reduce the impact of non-IID data on FL. However, due to the distributed nature, it is challenging for FL to analyze the contribution of clients with the absence of their data.

To address the above issue, this paper aims to propose an efficient federated weight determination mechanism for FL. By learning the characteristics of global model parameters and local model parameters, we use Deep Reinforcement Learning (DRL) method to dynamically evaluate the contribution of clients. Based on this, an adaptive weight is assigned for each client, which is used in global model aggregation.

2 PRELIMINARIES AND PROBLEM FORMULATION

2.1 Federated learning

We first introduce FL briefly. FL is a promising distributed machine learning, which trains a shared global model by aggregating clients' model updates at a central server. In particular, clients only need to upload their local models to the server in each communication round, without sharing their privacy-sensitive data. A typical training process of FL needs the the collaboration between clients and FL server. Specifically, the server first initiates the training task and send the global model parameters to the clients. The clients use their own data to train the model, and upload their training models to the server. After that, the server aggregates the global model, and sends it to clients for further training. The above process continues until a desirable training accuracy is achieved or the global loss function converges.

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

FL can be used in many machine learning models, aiming at minimizing the loss function of the training task. Suppose that the loss function of data sample k can be defined as $f_k(\boldsymbol{\omega})$, where $\boldsymbol{\omega}$ is the parameter vector of the trained model. Then, the loss function of client i on dataset D_i, denoted by $F_i(\boldsymbol{\omega})$, can be expressed as follows,

$$F_i(\boldsymbol{\omega}) = \frac{1}{|D_i|}\Sigma_{k\in D_i} f_k(\boldsymbol{\omega}). \tag{1}$$

Correspondingly, for a specific machine learning task, the global loss function of FL can be formulated into

$$F(\boldsymbol{\omega}) = \sum_{i=1}^{N} \frac{|D_i|F_i(\boldsymbol{\omega})}{\mathcal{D}}, \tag{2}$$

where $\mathcal{D}$ denotes the overall size of the datasets at clients, i.e., $\mathcal{D} = \sum_{i=1}^{N} |D_i|$.

2.2 Problem Formulation

This paper considers a general FL system, consisting of a central server and a set of N clients in $\mathcal{N}$. Each client i has a local dataset D_i, and its size is $|D_i|$. Due to the heterogeneous applications of devices, the data owned by clients are usually with different classes and amounts, appearing different non-IID degrees. In this context, the local models from different clients may have different contributions in global model aggregation. Most of existing works simply differentiate the contributions of local models using the aggregation weight, which is defined as the amount of local data over the size of all the data. However, the relationship between model accuracy and the amount of training data is nonlinear [12, 13], especially when the data is non-IID.

Intuitively, if the aggregation weight is set to be proportional to the contribution of client, the performance of FL can be further improved. Inspired by the above discussions, this paper considers a novel aggregation weight assignment problem, making sure that global model aggregation in FL can accurately capture the heterogeneous contributions of clients. Considering the above challenges, there are the following two problems to be addressed:

- **Contribution evaluation**: how to differentiate the contributions of clients, with the absence of data at the server?
- **Aggregation weight determination**: knowing the contribution of clients, how to assign appropriate weight for each client in global model aggregation?

To address the above issues, this paper proposes a novel DRL-based FL algorithm. By learning the characteristics of global model parameters and local model parameters, the proposed DRL dynamically identifies the contributions of clients in each communication round, based on which, the aggregation weights are appropriate assigned to the clients. To ease presentation, we refer to our aggregation Weight Assignment problem in Federated Learning as *FedWA* problem.

Without loss of generality, we use φ_t^i to denote the aggregation weight assigned to client i in communication round t. Then, the global model loss function can then be aggregated as follows,

$$F(\boldsymbol{\omega}) = \sum_{i=1}^{N} \varphi_t^i F_i(\boldsymbol{\omega}), \tag{3}$$

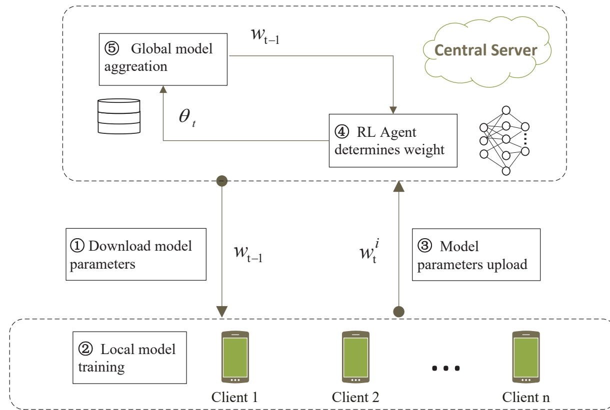

Figure 1: The training process of FedWA

which needs to be minimized by choosing appropriate aggregation weight φ_t^i.

3 THE FRAMEWORK OF FEDWA

We present the framework of the proposed FedWA as fllows. Similar to previous works, the whole FL training is conducted round by round. As shown in Fig. 1, the framework of FedWA can be described as follows.

- First, the server initiates the training task. And in each communication round, the server send global model parameters to clients.
- The participating clients locally train the model based on their own data, and upload the training models to the server.
- Using the received local models and current global model, the RL agent at the server utilizes DRL to learn the contribution of each client. Based on this, the server decides the aggregation weight of each client.
- The server aggregates the global model using the determined aggregation weight of each client, and send the new global model to the clients for the next communication round.
- Continue the above process, until reaching the desirable training accuracy or finishing the predefined communication rounds.

Therefore, the defined FedWA mainly focuses on learning the contributions of clients in each communication round, and determining the aggregation weights at FL server.

Finally, extensive simulations are conducted to confirm the effectiveness of the proposed mechanism in improving the training accuracy, and reducing the number of communication rounds.

4 ACKNOWLEDGEMENT

This work is supported by the National Key Research and Development Program of China under grant No. 2019YFB2102200, in part by National Natural Science Foundation of China under Grant No. 61872150. It was also supported in part by Guangdong Basic and Applied Basic Research Foundation under Grant No. 2020A1515011209, the Science and Technology Program of Guangzhou under Grant No. 202002030426. (Corresponding author: Xiumin Wang.)

REFERENCES

[1] M. Chiang and T. Zhang, “Fog and IoT: an overview of research opportunities,” *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, Dec. 2016.

[2] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proceedings of Machine Learning Research*, vol. 54, pp. 1273–1282, Apr. 2017.

[3] B. McMahan and D. Ramage, “Federated learning: collaborative machine learning without centralized training data,” Apr. 2017, [online] Available: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html.

[4] J. Konecny, B. McMahan, and D. Ramage, “Federated optimization: distributed optimization beyond the datacenter,” arXiv preprint arXiv:1511.03575, 2015.

[5] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen and M. Chen, “In-Edge AI: intelligentizing mobile edge computing, caching and communication by federated learning,” *IEEE Network*, vol. 33, no. 5, pp. 156–165, 2019.

[6] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, “Federated learning with non-IID data,” arXiv preprint arXiv:1806.00582, 2018.

[7] M. Duan, D. Liu, X. Chen, Y. Tan, J. Ren, L. Qiao and L. Liang, “Astraea: self-balancing federated learning for improving classification accuracy of mobile deep learning applications,” in *Proceedings of 2019 IEEE 37th International Conference on Computer Design* (ICCD), Abu Dhabi, United Arab Emirates, pp. 246–254, 2019.

[8] W. Zhang, X. Wang, P. Zhou, W. Wu and X. Zhang, "Client Selection for Federated Learning With Non-IID Data in Mobile Edge Computing," in *IEEE Access* , vol. 9, pp. 24462-24474, 2021, doi: 10.1109/ACCESS.2021.3056919.

[9] A. Hard, K. Rao, R. Mathews, F. Beaufays, S. Augenstein, H. Eichner, C. Kiddon, and D. Ramage, “Federated learning for mobile keyboard prediction,” arXiv preprint arXiv:1811.03604, 2018.

[10] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, “Federated learning of predictive models from federated electronic health records,” *International journal of medical informatics*, vol. 112, pp. 59–67, 2018.

[11] Jie Xu, Benjamin S Glicksberg, Chang Su, Peter Walker, Jiang Bian, and Fei Wang. Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research*, pages 1-19, 2020.

[12] Y. Peng, Y. Bao, Y. Chen, C. Wu, and C. Guo, “Optimus: An efficient dynamic resource scheduler for deep learning clusters,” in *Proc. ACM EuroSys*, 2018, pp. 1–14.

[13] Zhan Y, Li P, Wang K, et al. Big Data Analytics by CrowdLearning: Architecture and Mechanism Design[J]. *IEEE Network*, 2020, PP(99):1-5.