IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

DOI: 10.17148/IJARCCE.2023.12919

A DYNAMIC RESOURCE ALLOCATION FOR HIERARCHICAL FEDERATED LEARNING USING DECENTRALIZED EDGE INTELLIGENCE

Harish Babu P[1], Sundar Rajan[2], Kumaran. M[3]

Student ME (CSE), Jaya Engineering College, Thiruninravur, Thiruvallur, TamilNadu[1]

Professor, Department of (CSE), Jaya Engineering College, Thiruninravur, Thiruvallur, TamilNadu[2]

Professor, Department of (CSE), Jaya Engineering College, Thiruninravur, Thiruvallur, TamilNadu[3]

Abstract: To enable the large scale and efficient deployment of Artificial Intelligence (AI), the confluence of AI and Edge Computing has given rise to Edge Intelligence, which leverages on the computation and communication capabilities of end devices and edge servers to process data closer to where it is produced. One of the enabling technologies of Edge Intelligence is the privacy preserving machine learning paradigm known as Federated Learning (FL), which enables data owners to conduct model training without having to transmit their raw data to third-party servers. However, the FL network is envisioned to involve thousands of heterogeneous distributed devices. As a result, communication inefficiency remains a key bottleneck

Keywords: cloud, network, infrastructure, data security

I. INTRODUCTION

Today, the predominant approach for Artificial Intelligence (AI) based model training is cloud-centric, i.e., the data owners transmit the training data to a public cloud server for processing. However, this is no longer desirable due to the following reasons. Firstly, privacy laws, e.g., the General Data Protection Regulation (GDPR) [1], are increasingly stringent. In addition, the privacy-sensitive data owners can opt out of data sharing with third parties. Secondly, the transfer of massive quantities of data to the distant cloud burdens the communication networks and incurs unacceptable latency especially for time-sensitive tasks. As such, this necessitates the proposal of Edge Computing [2] as an alternative, in which raw data are processed at the edge of the network, closer to where data are produced. The confluence of Edge Computing and AI gives rise to Edge Intelligence, which leverages on the storage, communication, and computation capabilities of end devices and edge servers to enable edge caching, model training, and inference [3] closer to where data are produced. One of the enabling technologies [4] of Edge Intelligence is the privacy preserving machine learning paradigm termed Federated Learning (FL) [5]. In FL, only the updated model parameters, rather than the raw data, need to be transmitted back to the model owner for global aggregation.

The main advantages of FL are: (i) FL enables privacy preserving collaborative machine learning, (ii) FL leverages on the computation capabilities of IoT devices for local model training, thus reducing the computation workload of the cloud, and (iii) Model parameters are often smaller in size than raw data, thus alleviating the burden on backbone communication networks. This has enabled several practical applications, e.g., in the development of next-word-prediction models for text messaging [6], healthcare [7], Unmanned Aerial Vehicles (UAV) sensing [8], and mobile edge computing [9]. However, the FL network is envisioned to involve thousands of heterogeneous distributed devices, e.g., smartphones and Internet of Thing (IoT) devices [10]. In this case, the communication inefficiency remains a key bottleneck in FL. Specifically, node failures and device dropouts due to communication failures can lead to inefficient FL. Moreover, workers, i.e., data owners, with severely limited connectivity are unable to participate in the FL training, thus adversely affecting the model's ability to generalize. As such, solutions from edge computing have recently been incorporated to solve the communication bottleneck in FL. In [4], [11], [12], a hierarchical FL (HFL) framework is proposed in which the workers do not communicate directly with a central controller, i.e., the model owner. Instead, the local parameter values are first uploaded to edge servers, e.g., at base stations, for intermediate aggregation. Then, communication with the model owner is further established for global aggregation.

Besides reducing the instances of global communications with the remote servers of the model owner, this relay approach reduces the dropout rate of devices. While [11] discusses convergence guarantees and presents empirical results to show that the HFL approach does not compromise on model performance, the challenges of resource allocation and incentive mechanism design have not yet been well-addressed in the HFL framework. In 5G and Beyond networks, the resource sharing and incentive mechanism design for end-edge-cloud collaboration is of paramount importance to facilitate efficient Edge Intelligence [4].

In this project, we consider a decentralized learning based system model inspired by the HFL. In our system model, there exist data owners, hereinafter referred to as workers, that participate in the FL model training facilitated by different cluster heads, e.g., base stations that support the intermediate aggregation of model parameters and efficient relaying to the model owners (Fig. 1). We consider a two level resource allocation and incentive design problem as follows:

1) Lower level (Between workers and cluster heads): Each worker can freely choose which cluster to join. To encourage the participation of workers, the cluster heads offer reward pools to be shared among workers based on their data contribution in the cluster. For example, a worker that has contributed more data during its local training will receive a larger share of the reward pool. Moreover, the cluster heads offer the workers resource blocks, i.e., bandwidth, to facilitate efficient uplink transmission of the updated model parameters. However, as more workers join a cluster, the payoffs are inevitably reduced due to the division of rewards over a larger number of workers and the increased communication congestion.

Thus, the cluster selection strategies of each worker can affect the payoffs of other workers. Accordingly, the workers may slowly adapt their strategies in response to other workers. In contrast to conventional optimization approaches, we use the evolutionary game theory [14] to derive the equilibrium composition of the clusters. Our game formulation enables the bounded rationality and worker dynamics to be captured. Specifically, the workers gradually adapt their strategies in response to other non-cooperative workers. To achieve their objectives, they observe each others' strategies and gradually adjust their strategies accordingly.

2) Upper level (Between cluster heads and model owners):

There may be multiple model owners in the network that aim to train a model for their respective usage collaboratively with the participation of the workers and cluster heads. However, at any point of time, each worker and cluster head can only participate in the training process with a single model owner.

To derive the allocation of cluster head to the model owner, as well as the optimal pricing of the services of the cluster head by the competitive model owners, we adopt a deep learning based auction mechanism which preserves the properties of truthfulness of the bidders, while simultaneously achieving revenue maximization for the cluster heads.

The main contributions of our project are as follows:

1) We propose a joint resource allocation and incentive design framework for the HFL. The "Edge for AI" [15] approach supports decentralized Edge Intelligence, i.e., FL at the edge with reduced reliance on a central controller.

2) We model the cluster selection decisions of the workers as an evolutionary game. Then, we provide proofs for the uniqueness and stability of the evolutionary equilibrium. In contrast to conventional optimization tools which assume that the players are perfectly rational, our model enables us to capture the dynamics and bounded rationality of player decisions.

3) To assign the cluster heads to model owners, we use a deep learning based auction mechanism. In contrast to conventional auctions, the deep learning based auction ensures seller revenue maximization while satisfying the individual rationality and incentive compatibility constraints.

1. SYSTEM STUDY

1.1 Existing System Int

The existing server and cloud infrastructure are not capable of handling. Some application requires low latency or real-time results the existing architectures are not capable of it. One more issue is the privacy of data, as users have to share data with servers.

DOI: 10.17148/IJARCCE.2023.12919

1.2 Literature Survey

The literatures [8–10] is biased towards specific areas, focusing on the statistical challenges, communication efficiency/stability issues, device/data heterogeneity issues, privacy and security issues, traceability and accountability in the process of combining mobile edge computing, healthcare and FL. They also give current solutions. The literature [12] discusses several possible applications in 5G networks: edge computing and caching, spectrum management, 5G core network, and described the key technical challenges for future research of FL in the wireless communication environment: security and privacy challenges, algorithms Related challenges, challenges in the wireless environment. •

The literatures [7,11,13,14] discuss data privacy protection in FL. Literature [7] discusses valuable attack mechanisms and proposes corresponding solutions to corresponding attacks. Literature [11] introduces the basic knowledge and key technologies of various attacks and discusses the future research direction of achieving more robust privacy protection in FL. The literature [13] discusses the protection of privacy and security when designing the FL system and divides the methods into three categories: client-side privacy protection, server-side privacy protection, and security protection for FL.

The privacy and security issues of FL are divided into convergence, data poisoning, scaling up, and model aggregation problems. The author provides solutions to related problems. Literature [14] analyzes the federated security problems under Byzantine adversaries and divides the current security FL algorithms (SFLAs) into four categories: aggregation rule-based SFLAs, preprocessing based SFLAs, model-based SFLAs, and adversarial detection-based SFLAs.

The author provided a qualitative comparison of current SFLAs and reviewed some typical work on SDLAs. • The literature [5]briefly reviews semi-supervised algorithms and FL, and then analyzed federated semi-supervised learning, including settings and potential methods.

1.3 Proposed System

The proposed system aims a resource allocation and incentive mechanism design framework for HFL. We considered a two-level problem and leveraged on the evolutionary game theory to derive the equilibrium solution for the cluster selection phase. Then, we introduced a deep learning based auction mechanism to value the cluster head’s services.

The performance evaluation shows the uniqueness and stability of the evolutionary equilibrium, as well as the revenue maximizing property of the auction mechanism. In the future work, we will consider social network effects and their impact on the cluster selection decisions of the workers, as in [26]. Moreover, we may also account for the existence of malicious workers.

1.4 ORGANIZATION OF THE REPORT

The chapter 1 presents review of literature. The goal tended to be attained in the project is explained in objectives. The problem description tells the need for the system with the advantage of proposed system over existing system.

The chapter 2 explains the system requirement for both, by feature and by functionality hierarchy.

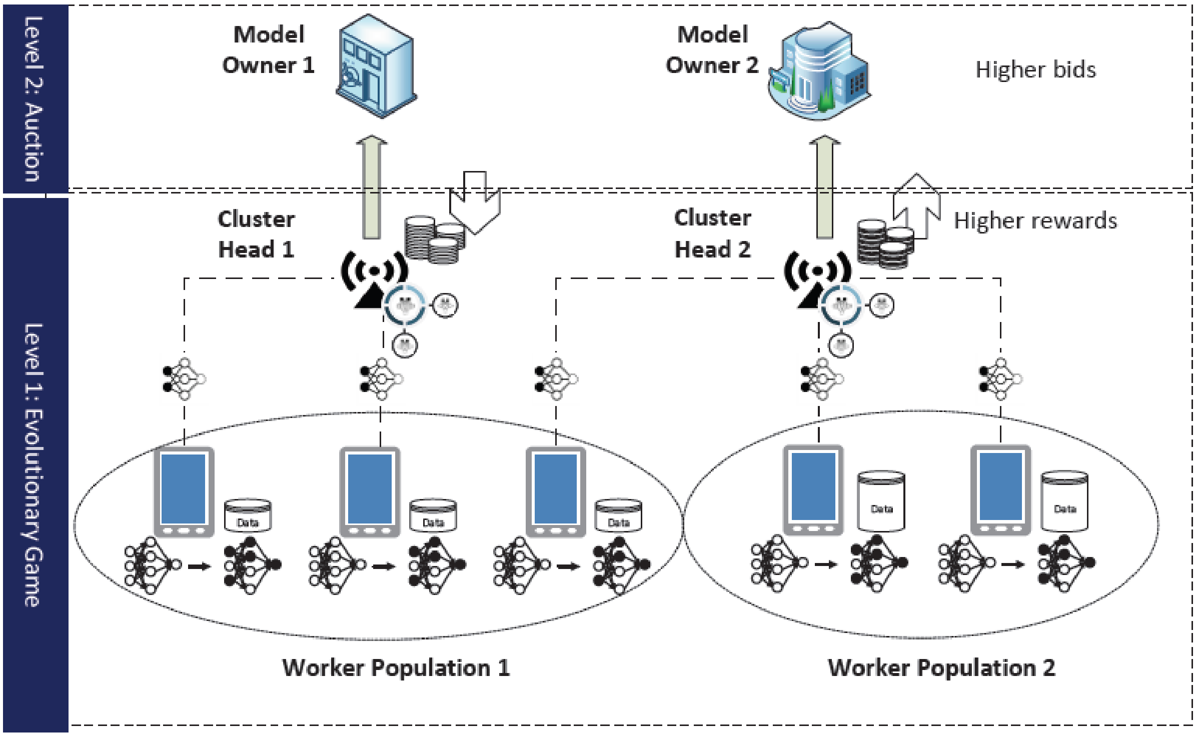
The chapter 3 describes the system design which includes the decomposition description, dependency description and detailed design of modules.

The chapter 4 describes the implementation of the project, which includes the modules and components used in this project.

The chapter 5 deals with the test plan and testing of the project. The chapter 6 describes the result of the implementation. The chapter 7 contains the conclusion of the work done and also the extension of the work.

DOI: 10.17148/IJARCCE.2023.12919

2. DESIGN AND IMPLEMENTATION CONSTRAINTS



2.1 SYSTEM FEATURES

1) Local Computation: Each worker trains the received global model w(k) locally.
2) Wireless Transmission: The worker transmits the model parameter update to its cluster head j.
3) Intermediate Model Parameter Update: All parameter updates received from its pjn workers are aggregated by the cluster head to derive an updated intermediate model w(k+1) j , which is then transmitted back to the worker for the (k+1)th training iteration.

2.2 OTHER NON FUNCTIONAL REQUIREMENTS

2.2.1 Performance Requirements

- The major aim for choosing the domain of Data Analysis is for the velocity criteria of data processing.
- Connecting the commodity systems and forming the nodes between them helps in quick retrieval of data items.
- It is verified that the product requirements are being met to help and plan resource requirements.

2.2.2 Safety Requirements

- The protection of computer based resources that includes Hardware, Software, Data, Procedures and people against unauthorized use or natural.
- Data Processing large datasets which need higher storage capacity than an allocated Data node may lead to failure of data processing and failure of that node.
- The developer of the application need to write complex business logic in order to successfully execute the join queries.

2.2.3 Security Requirements

- Ensuring the proper authentication of users who access FL.
- Ensuring that data access histories for all users are recorded in accordance with compliance regulations and for other important purposes.
- Ensuring the protection of data both at rest and in transit through enterprise-grade encryption.

DOI: 10.17148/IJARCCE.2023.12919

II. CONCLUSION

In this project, we proposed a resource allocation and incentive mechanism design framework for HFL. We considered a two-level problem and leveraged on the evolutionary game theory to derive the equilibrium solution for the cluster selection phase. Then, we introduced a deep learning based auction mechanism to value the cluster head's services. The performance evaluation shows the uniqueness and stability of the evolutionary equilibrium, as well as the revenue maximizing property of the auction mechanism. In the future work, we will consider social network effects and their impact on the cluster selection decisions of the workers, as in Moreover, we may also account for the existence of malicious workers.

REFERENCES

[1]. W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," I, 2019.

[2]. W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE internet of things journal, vol. 3, no. 5, pp. 637–646, 2016.

[3]. D. Xu, T. Li, Y. Li, X. Su, S. Tarkoma, and P. Hui, "A survey on edge intelligence," arXiv preprint arXiv:2003.12172, 2020.

[4]. W. Y. B. Lim, J. S. Ng, Z. Xiong, D. Niyato, C. Leung, C. Miao, and Q. Yang, "Incentive mechanism design for resource sharing in collaborative edge learning," arXiv preprint arXiv:2006.00511, 2020.

[5]. H. B. McMahan, E. Moore, D. Ramage, S. Hampson et al., "Communication-efficient learning of deep networks from decentralized data," arXiv preprint arXiv:1602.05629, 2016.