Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 202 (2022) 348–353

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Identification, Information and Knowledge in the internet of Things, 2021

Blockchain Empowered Federated Learning for Data Sharing Incentive Mechanism

Zexin Wang[a], Biwei Yan[a], Anming Dong[a,*]

[a] *School of Computer Science and Technology, Qilu University of Technology(Shandong Academy of Science), Jinan and 250300, China*

Abstract

In the machine learning, data sharing between different participants can increase the amount of data, improve the quality of the dataset, and thereby improve the quality of the model. Under the condition of data supervision, federated learning, as a distributed machine learning, aims to protect data while training models through collaboration among all parties to achieve data sharing and improve model quality. However, there are still some issues. For instance, the lack of trust between the participants makes it impossible to establish a secure and reliable sharing mechanism. In addition, how to fairly share the benefits generated by the model, identify honest participants and punish malicious participants is still a challenge. In this paper, we propose a new federated learning scheme based on blockchain architecture for federated learning data sharing. Moreover, an incentive mechanism based on reputation points and Shaply values is proposed to improve the sustainability of the federated learning system, which provides a credible participation mechanism for data sharing based on federated learning and fair incentives. The experimental results and analysis show that the loss of federated learning is more smooth than that of centralized machine learning.

Keywords: Federated learning; Blockchain,Incentive Mechanism; Shaply Value;

1. Introduction

With the development of machine learning, the application of artificial intelligence in many fields is advancing by leaps and bounds. However, due to the competitive relationship between the parties, it is difficult to share the data of the parties, resulting in a state of fragmentation of the data, and restricting the development of the artificial intelligence industry that relies heavily on data[1]. Large datasets play a very important role in the development of machine learning. The data that can be obtained in real life is either small in scale or low in quality. Therefore, it is very difficult to obtain a large amount of high-quality data. Meanwhile, people's attention to data privacy and security

* Corresponding author. Tel.: +1-516-885-5183.
E-mail address: anmingdong@qlu.edu.cn

10.1016/j.procs.2022.04.047

is also increasing. Many legislative bodies and regulatory agencies have issued laws to regulate the management and use of data. In 2018, the General Data Protection Regulation (GDPR) is implemented by the European Union.

In this environment, federated learning was proposed by McMahan in 2016[2]. The core idea of federated learning is to train machine learning models without having to store all the data centrally. Each user with data resources can use their own data to train a model separately, and then through model sharing, a global model is obtained through model aggregation, and privacy protection technology is used to reduce the risk of data leakage.

Federated learning can share private local data under the premise of protecting data privacy, and collaboratively build and improve the quality of machine learning models. Data is the "raw material" for training models. By incentivizing data owners to contribute high-quality "raw materials", the training model can be more efficiently trained. However, in real scenarios, how to reward model contributors so as to encourage more participants to join the federated system and stimulate them to share high-quality data for a long time has become a problem to be solved in federated learning.

The main contributions of this paper is as follows:

1. In order to avoid a single point of failure in the aggregation of federated learning centers, we propose a blockchain-based federated learning sharing model. The decentralization of the blockchain can avoid the single point of failure of the federated learning center aggregation, and the blockchain cannot be tampered with. The characteristics of this can prevent the local model parameters from being tampered with and affecting the overall model performance.

2. Incentives are an important factor in maintaining data sharing among participants. We propose an incentive mechanism based on reputation points and shaply values. Participants who contribute to the model will be rewarded to ensure the sustainable development of federated learning.

3. In order to ensure the fairness of incentives, we have designed a smart contract for the automatic distribution of profits on the blockchain that avoids the participation of third parties.

The rest of the paper is organized as follows: Section 2 shows the preliminaries. Section 3 introduces our data sharing model and designs the incentive mechanism. Section 4 gives the security analysis and performance evaluation. Finally, the paper gives a summary of the paper in Section 5.

2. Related Work

Incentive mechanism design is an important research direction in the field of federated learning [3][4].We reviewed existing work on federated learning incentive design. At present, the most widely adopted method to fairly evaluate the contributions of federated learning participants is the shapley value method[5][6]. Liu et al. propose FedCoin, which is a blockchain-based peer-to-peer payment system for FL to enable a feasible SV based profit distribution. In FedCoin, blockchain consensus entities calculate SVs and a new block is created based on the proof of Shapley (PoSap) protocol.[7] Zhao et al. designed an incentive mechanism to prevent the poisoning attack as well as reward contributors properly by combining the Multi-KRUM and the reputation-based incentive protocols[8]. Lim et al.[10] proposed a hierarchical incentive mechanism framework. Using the backward induction, that first solve the contract formulation and then proceed to solve the coalitional game with the merge and split algorithm. Sharma et al. present a platform architecture of blockchain-based federated learning systems for failure detection in IIoT. In addition, to motivate clients to participate in federated learning, a smart contact-based incentive mechanism is designed depending on the size and the centroid distance of client data used in local model training[11].

BlockFL [9] uses the blockchain to reward users for local updates. The reward is proportional to how many local data points are used. However, this value may be exaggerated by malicious nodes seeking higher returns. The paper[12] proposed an effective incentive mechanism that combined reputation and contract theory to encourage high-reputation mobile devices with high-quality data to participate in model learning. Song et al. proposed the contribution index, a new Shapley value based metric fit for assessing the contribution of each data provider for the joint model trained by federated learning[6].

Incentive mechanisms are usually related to mathematical knowledge such as game theory. Here, we also review the work of existing revenue-sharing games. The paper [13] proposed three profit sharing schemes based on marginal contribution, and all three profit distribution schemes recognize Nash equilibrium.

3. System model

In this paper, we consider a decentralized and private multi-party data sharing scenario. In order for high quality and continuous data owners to participate in enduring data sharing, appropriate incentives are necessary. Here, we first introduce our proposed data sharing model based on blockchain and federated learning, and then describe the incentive mechanism based on reputation points and shaply values.

3.1. Blockchain empowered federated learning data sharing model

The sharing model is divided into two parts, the federated learning network and the blockchain network. The nodes participating in the network can be divided into four categories: federated learning requester, federated model participants, computing center, aggregation node, smart contract, as shown in the Figure 3.

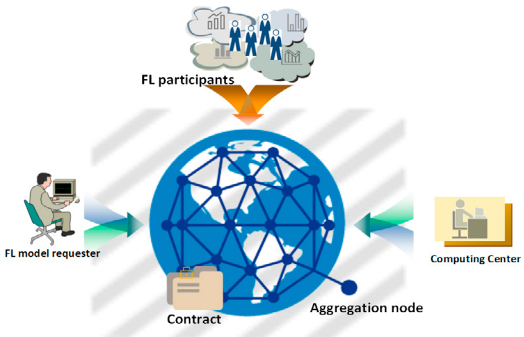

Fig. 1. Data sharing model

The federated learning network includes data requester and data participants. The data demanders can submit data sharing requests to the blockchain. The data owners who match the needs train the local model locally and upload the local model to the blockchain perform model aggregation and update. In the blockchain network, the identity ID and data summary of each participant are recorded in the blockchain network in the form of a transaction, and the data sharing event is also stored as a transaction in the blockchain and verified in the form of a merkle tree.The incentive mechanism based on the blockchain network is an important factor for more participants to join the network, and the automatic distribution of benefits by smart contracts is a guarantee of fairness. The blockchain guarantees that the model sharing process is safe, and can identify malicious participants and protect the quality of the overall model.

The nodes participating in the network can be divided into the following five categories:

- Federated Learning Requester: refers to the user node that needs to train a federated learning model. The demander publishes tasks and rewards to the blockchain network, and multiple data owners train the federated learning model to obtain rewards.
- Federated Learning Participants: Participants hold distributed data and get paid by cooperating to complete the tasks of the demand side. Participants train a local model based on local data and upload the model to the blockchain for model aggregation. After the task is completed, the blockchain smart contract will automatically distribute the benefits.
- Aggregation Node: It is a node on the blockchain. Miners get rewards by contributing their own computing power aggregation model. Only the node that completes the aggregation model first can get rewards in each round.
- Computing Center: As the number of nodes grows, the calculation of shaply value is a huge problem. We outsource the calculation of shaply value to the computing center to reduce the amount of calculation for blockchain nodes.

- Smart Contracts: Smart contracts are automatically executed codes deployed on the blockchain to manage revenue distribution and payment standards. Once the data demander pays for the demand, the smart contract will allocate funds according to a predetermined plan to ensure fairness.

3.2. Incentive mechanism

Although the contribution of participants to the federated model is an important consideration, it is not the only criterion. In order to ensure that participants can share data honestly, we designed a dual incentive scheme based on reputation points and shaply values. Each parameter holder initializes a reputation value when registering. With each round of data sharing activities, honest participants will be rewarded with reputation value and revenue, and malicious participants who share false parameters will be punished and reduced reputation points. When the reputation score is lower than the threshold, participation in data sharing activities will be prohibited to punish malicious participants for damaging the model.

The total income P is divided into the following parts, and the specific ratio can be agreed in advance through smart contracts:

- Rewards paid to data providers
- The cost of processing model aggregation paid to the aggregation node
- Calculation shaply remuneration paid to the computing center

The shaply value is most often accompanied by collaboration and sharing scenarios, so we choose the shaply value as the allocation method. Shaply value is a measure based on the marginal contribution of the impact of participants joining the collective in different orders on the collective benefits, which can more fairly estimate the contribution of participants to the collective. However, the contribution value is not the only measurement factor. In order to prevent malicious participants from sharing false parameters to damage the model, we introduce reputation points to judge the honesty of participants.

We assume that the contribution of the data provider is the marginal optimization of the loss function of the model, and we choose the loss function as an indicator of the value function. In different scenarios, the value function can also be replaced with other reasonable indicators to adapt to more sharing scenarios.

The shaply value of node i is denoted as φ_i, assuming that the number of participants in a round of data collaboration training is $\mathcal{N}$, and $N = |\mathcal{N}|$ represents the number of participants in the group. We call a participant a federated learning node, and S represents the possible alliance of N nodes, so S is a subset of N, denoted as $S \subseteq N$. Participants' model optimization of alliance $S \subseteq N \setminus \{i\}$ is defined as:

$$\Delta i(\mathcal{F}, S) = \mathcal{F}_s(\omega) - \mathcal{F}_{s\cup i}(\omega) \tag{1}$$

The relationship between shaply and loss function can be defined as follows:

$$\varphi_i(N, \mathcal{F}) = \sum_{s\in S_i} w(|s|)\frac{\mathcal{F}_s(\omega)-\mathcal{F}_{s\cup i}(\omega)}{\mathcal{F}_N(\omega)} v(P)) \tag{2}$$

Where $f_i(N) = \sum_{s\in S_i} w(|s|)\frac{\mathcal{F}_s(\omega)-\mathcal{F}_{s\cup i}(\omega)}{\mathcal{F}_N(\omega)}$is the influence factor of participant i , $v(P)$ is the reward paid to the data provider in the total revenue. We quantify the participant's contribution by $\Delta i(\mathcal{F}, S) = \mathcal{F}_s(\omega) - \mathcal{F}_{s\cup i}(\omega)$,and normalize the marginal contribution by dividing by the maximum value of $\mathcal{F}_N(\omega)$.

The reputation reward of participant i consists of reputation points and data value. We suppose that the reputation score in the t-th sharing process can be expressed as a vector $R^t_j = \left\{b^t_j, d^t_j\right\}$. Among them, b^t_j is the positive integral, and d^t_j is the negative integral.According to the reputation score tuple vector, the reputation score of the participant participating in the sharing activity for the t-th time is: $T^t_i = b^t_i + \alpha b^t_i$ Where α is the negative integral influence coefficient.

There are many factors that affect the value of data. We believe that the value of data changes over time:$W_i(t) = W_0 Q^{-\theta t}$.where $Q \in (0, 1)$ is the given change parameter,$\theta > 0$ is the exponential decay constant.Therefore, the reputation of participant i at the t-th time is expressed as:$Z^t_i = T^t_i + W^t_i$.

The incentive obtained by participant i in the t-th sharing is: $M_i^t = \varphi_i^t v(P) + Z_i^t c(P)$.The total incentive of the data provider is expressed as

$$M_N^t = \sum_{i=1}^{N} M_i^t = \sum_{i=1}^{N} \varphi_i^t v(P) + Z_i^t c(P)) \tag{3}$$

Each round of honest participants will get certain positive points, and the false parameters shared by malicious participants will damage the model and increase negative points. Each round of sharing will update the reputation points of the participants. Credit scores below the threshold will be restricted from participating in sharing activities, so the cost of nodes doing evil is very high, so as to avoid malicious participants from destroying the model.

4. Safety analysis and experiment

4.1. Safety analysis

Data security: Due to the reproducibility of data, it is difficult to share data information without leaking the data. The real data is stored locally in the participants, and only the local data is trained.

Remove centralized trust: Our sharing model designs distributed blockchain nodes to replace the federated learning Center server for model aggregation, removing the high risk of data leakage caused by the federated learning center aggregation model.

Shared data quality: A dual incentive mechanism based on reputation and shaply value is proposed. The reputation mechanism can ensure the honest participation of data owners in sharing, and shaply value encourages participants to contribute high-quality data. In addition, the smart contract in the blockchain can detect the quality of the participant's local model and prevent malicious participants from using low-quality parameters to destroy the effect of the aggregation model.

4.2. Experiment

In this part, we use Python 3.7 to test the comparison of the loss function value of the two training methods of federated learning and centralized learning. The model uses the built-in ResNet-18 model of torchvision. In Figure 5, we can see that the training effect of federated learning is similar to that of centralized learning, but the loss value of federated learning decreases more smoothly.

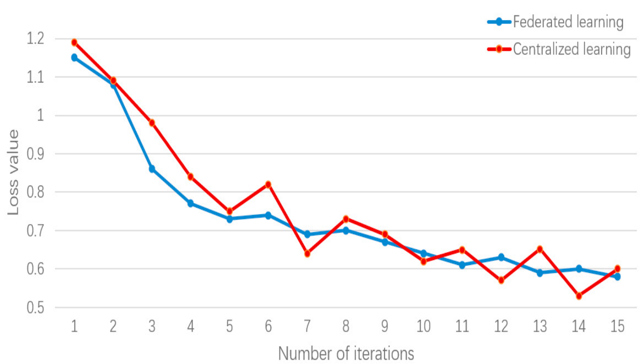

Fig. 2. Comparison of two different training methods

5. Conclusion

Combining federated learning and blockchain technology, a new data sharing model is proposed, and an incentive mechanism based on reputation points and shaply values is designed. The main purpose of the combination of federated learning and blockchain technology is to provide a credible participation environment and automatic revenue sharing mechanism. Considering that malicious nodes will destroy the global model parameters, this incentive mechanism combines reputation points, and low-point nodes are judged as malicious nodes to restrict participation in sharing and avoid the destruction of the global model. Federated incentives can allow more hungry participants to join the federal ecology. The increase in the amount of data will help improve the effect of the model, and also allow participants to obtain more benefits and achieve a win-win situation.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (NSFC) under Grant 61832012, 61771289 and 61672321, the Key Research and Development Program of Shandong Province under Grant 2019JZZY020124 and the Pilot Project for Integrated Innovation of Science, Education and Industry of Qilu University of Technology (Shandong Academy of Sciences) under Grant 2020KJCZD02.

References

[1] Yang, Q., Liu, Y., Chen, T., and Tong, Y. (2019). "Federated machine learning: Concept and applications". *ACM Transactions on Intelligent Systems and Technology (TIST)* 10(2), 1-19.

[2] McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017, April). "ommunication-efficient learning of deep networks from decentralized data." *In Artificial intelligence and statistics* (pp. 1273-1282). PMLR.

[3] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N.,Bonawitz, K.,Charles, Z.,Cormode, G.,Cummings, R.and other. (2019) "Advances and open problems in federated learning." *arXiv preprint arXiv* 1912.04977.

[4] Yang, Q., Liu, Y., Chen, T., and Tong, Y. (2019) "Federated machine learning: Concept and applications. " *ACM Transactions on Intelligent Systems and Technology(TIST)* **10** (2): 1–19.

[5] Jia, R., Dao, D., Wang, B., Hubis, F. A., Hynes, N., Gürel, N. M., ... and Spanos, C. J. (2019) "Towards efficient data valuation based on the shapley value. " *In The 22nd International Conference on Artificial Intelligence and Statistics* 1167–1176.

[6] Song, T., Tong, Y., and Wei, S.(2019)"Profit allocation for federated learning." *In 2019 IEEE International Conference on Big Data (Big Data)* 2577–2586

[7] Liu, Y., Ai, Z., Sun, S., Zhang, S., Liu, Z., and Yu, H. (2020) "Fedcoin: A peer-to-peer payment system for federated learning." *In Federated Learning* 125–138.

[8] Zhao, Y., Zhao, J., Jiang, L., Tan, R., Niyato, D., Li, Z., Lyu, L. and Liu, Y. (2020) "Privacy-preserving blockchain-based federated learning for IoT devices." *IEEE Internet of Things Journal* **8** (3): 1817–1829.

[9] Kim, H., Park, J., Bennis, M., and Kim, S. L. (2018) "On-device federated learning via blockchain and its latency analysis." *arXiv preprint arXiv* 1808.03949.

[10] Lim, W. Y. B., Xiong, Z., Miao, C., Niyato, D., Yang, Q., Leung, C., and Poor, H. V. (2020). "Hierarchical incentive mechanism design for federated machine learning in mobile networks."*IEEE Internet of Things Journal*, 7(10), 9575-9588.

[11] Sharma, P. K., Park, J. H., and Cho, K. (2020). "Blockchain and federated learning-based distributed computing defence framework for sustainable society." *Sustainable Cities and Society*, 59, 102220.

[12] Kang, J., Xiong, Z., Niyato, D., Xie, S., and Zhang, J. (2019) "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory." *IEEE Internet of Things Journal* **6** (6): 10700-10714.

[13] Gollapudi, S., Kollias, K., Panigrahi, D., and Pliatsika, V. (2017) "Profit sharing and efficiency in utility games. " *In 25th Annual European Symposium on Algorithms (ESA 2017)*,Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[14] McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. (2017) "Communication-efficient learning of deep networks from decentralized data. " *In Artificial intelligence and statistics* 1273–1282.

[15] McMahan, H. B., Moore, E., Ramage, D., and y Arcas, B. A. (2016) "Federated learning of deep networks using model averaging." *arXiv preprint arXiv:* 1602.05629.

[16] Konečnỳ, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., and Bacon, D. (2016) "Federated learning: Strategies for improving communication efficiency." *arXiv preprint arXiv:* 1610.05492.

[17] McMahan, H. B., Moore, E., Ramage, D., and y Arcas, B. A. (2016) "Federated learning of deep networks using model averaging." *arXiv preprint arXiv:* 1602.05629.

[18] Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J.(2021) "Dive into deep learning. " *rXiv preprint arXiv:* 2106.11342.