A Dual Metric Mechanism for Federated Learning Based On Data and Arithmetic Resources

WenXia Ge, ShaoHeng Wang

Abstract—**With the increasingly stringent data privacy protection regulations and the decentralisation of computing resources, federated learning has gradually become a current research hotspot. Currently federated learning can effectively solve the problem of data silos, but there are still several notable challenges in the current federated learning in specific application practices, including the lack of efficient and secure incentive mechanisms, the increasingly obvious drawbacks of the traditional centralised management model, and the training security problems caused by malicious users. This paper is dedicated to solving these problems by proposing a secure federated learning mechanism based on a dual incentive strategy of data and arithmetic resources, aiming to improve the performance and security of the system through innovative algorithms and framework design.**

This paper proposes a dual metric mechanism based on data resources and arithmetic resources. Contribution measurement is the antecedent problem of designing the federated learning incentive mechanism, this paper improves the data resources and arithmetic resources of the federated training client to be evaluated from multiple dimensions, and selects multiple indicators for specific measurements, in which the innovative introduction of task relevance indicators can more accurately match the client with the learning task, which further improves the resource utilisation rate and the efficiency of federated training.

Index Terms—Federated **Learning,Incentive Mechanisms, Data Measurement ,Computing Power Measurement**

I. Introduction

Since its introduction, the concept of federated learning has been recognised by academics and related industrial fields, and has been developing rapidly, but at the same time, it also faces a lot of tests in practical applications. Firstly, users are not very active in federation training. The global model in the federation training process is obtained from the local model aggregation, so the training of the user's local model directly determines the degree of accuracy of the federation training model, and the user participation in the training process not only needs to provide data resources, but also inevitably take up a lot of arithmetic resources, such as consuming power, taking up communication resources, computational resources and so on. These additional payments lead to a significant reduction in the motivation of users to participate in federated learning, and it is necessary to provide users with relevant incentives to motivate their participation in training. In addition, federated learning security needs to be improved. Security threats may be encountered in all stages of the federated learning process; in the local training stage, malicious users can perform data poisoning attacks by participating in training with low-quality data [1]; in the process of uploading parameters, malicious users can perform model poisoning attacks to upload maliciously-designed model parameters; Byzantine failures are also common in distributed learning, in which the parties' models may perform poorly in learning and also upload updates randomly. All of the above attacks can lead to unsatisfactory training results for federated learning models. Meanwhile, since there are also problems in federated learning such as gradient inversion, free-rider attack, membership inference attack, etc., which may lead to partial data leakage of users, posing a threat to their privacy and security, and due to the concern of privacy leakage, the willingness of the users to participate in federated training is even more greatly reduced. All these put forward new requirements for federated learning to design a perfect user incentive mechanism while continuously improving the security of federated learning.

Existing research on federated learning mainly focuses on performance enhancement such as improving model accuracy and shortening learning task completion time. Research on the security of federated learning and the design of incentive mechanisms is still relatively scarce and needs further attention and exploration.

II. Related Work

Federated learning is a popular research direction in the field of machine learning in recent years, and its development is gradually changing the way of data processing and model training, which has a good development prospect. However, at present, federated learning still has problems such as unbalanced distribution of client data and arithmetic resources, client selfishness, and low user participation. In order to solve these problems, this paper conducts research on federated learning incentive mechanism. In the design of the federated learning incentive mechanism, since the final training effect needs to be determined after the federated training iteration is completely finished, it is not possible to directly allocate rewards based on the training effect of each client, so it is an important part to determine the contribution value of each client before the training starts. The amount of contribution of federated training clients is one of the key factors in incentive determination, which means that contribution assessment is a key antecedent problem in federated learning incentives . Therefore, this chapter first addresses the client contribution assessment problem in the design of federated learning incentive mechanism.

To address the contribution assessment problem, Wang et

Manuscript received May 06, 2024.

WenXia Ge, School of Software, Tiangong university, Tianjin ,China.

ShaoHeng Wang, School of computer science and technology, Tiangong university, Tianjin ,China.

al[2] comprehensively consider parameters such as the total number of training samples and the level of privacy protection to express the estimation quality of data owners, and use differential privacy to protect the privacy of the true cost model and the local model, and propose a quality-aware incentive mechanism under dual privacy protection. Although the privacy issue, which is easily ignored in the federated learning process, is taken into account, only the data quality is included in the influence of the federated training contribution and the variability in terms of client-side arithmetic resources is not taken into account. If only the data resources are considered to allocate rewards, the federated training clients with good arithmetic resources do get the same rewards despite the fact that they are used for a shorter period of time when the data resources are the same, which will discourage the clients' participation. Literature [3] [4]also all consider only the impact of data volume and data quality on incentives, and do not consider that the difference in arithmetic resources affects federation training time and user participation motivation.

In terms of data quality metrics, Li et al [5] proposed a simple and effective method to assess data quality from the perspective of information entropy by distinguishing good data from bad data through perturbation entropy, but this method has some limitations for federated learning. Different training iterations of federated learning may correspond to different training tasks, and good or bad data quality is also closely related to the training tasks; if the user dataset has a high evaluation quality score but is not related to the training tasks, the dataset has no value for model training.

III. Data Resource Metrics

In the data quality assessment process, we introduce three roles, the data owner, the task publisher and the data assessor (FL server). The task publisher publishes a federated learning task, which is typically a machine learning training model and a set of high-quality sample data that is well suited for the task. Each data owner has its own local data set, and the data evaluator scores each data owner's local data set based on the sample data set. As shown in the figure below, the specific data quality assessment is divided into two modules.

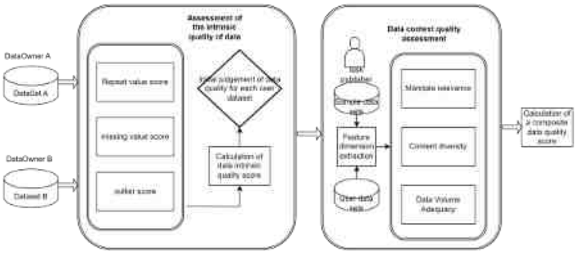
Assessment of the intrinsic quality of data
DataOwner A
DataSet A
DataOwner B
Dataset B
Repeat value score
missing value score
outlier score
Initial judgement of data quality for each user dataset
Calculation of data intrinsic quality score
Data context quality assessment
task publisher
Sample data sets
Feature dimension extraction
User data sets
Mandate relevance
Content diversity
Data Volume Adequacy
Calculation of a composite data quality score

Figure 3.1 Data quality assessment process

A. *Intrinsic Quality Assessment*

By evaluating the intrinsic quality dimensions, the data owner's dataset can be preliminarily screened, and the dataset that does not meet the training criteria can be judged as a poor-quality dataset, and only the dataset that is judged to meet the criteria can be further involved in the federated learning training.

The intrinsic quality of data is scored by four dimensions, which can effectively assess the accuracy and completeness of the dataset, including outlier scoring S_o , missing value scoring S_m and duplicate value scoring S_d .

Calculate outlier scores . Each data owner counts the outliers for each feature dimension of the local dataset, e.g., for discrete data, beyond the theoretical range of values of the data, some data values in the sample significantly deviate from the rest of the values of the sample points. Calculate the ratio of the number of samples whose features are outliers to the total number of samples, and calculate the outlier score based on the ratio :

$$S_o = \frac{1}{t}\sum_{i=1}^{t}\left(1-\frac{D_{O_i}}{D_A}\right) \tag{1}$$

Where, D_A is the total data samples of the user, D_{O_i} is the number of samples in which the feature in dimension i is an outlier. The higher the outlier score, the less outlier data appear in the local dataset.

Calculate missing scores. Each participating user processes the missing values of each feature dimension in the local dataset, assigns the missing value to "Null", and calculates the ratio of the total number of samples with the value type "Null" to the total number of samples for each feature dimension, using the formula:

$$S_m = \frac{1}{t}\sum_{i=1}^{t}\left(1-\frac{D_{M_i}}{D_A}\right) \tag{2}$$

Where D_{M_i} is the number of samples where the i-th dimension is characterised by missing values. The higher it is, the fewer missing values in the local data, and the higher the relative quality of the data.

Calculate the duplicate value . Each participant counts the number of duplicates in the local dataset and calculates the ratio of the number of duplicates to the total number of samples:

$$S_d = 1-\frac{D_D}{D_A} \tag{3}$$

Where, S_d is the number of repeated samples in the local dataset. The higher it is, the less local data and duplicates the user has, and the relatively more data information it contains.

The intrinsic quality score of each participant's dataset was calculated based on the scores of the above metrics:

$$S_I = S_d + S_m + S_o \tag{4}$$

B. *Contextual quality assessment*

Different federated training tasks have different needs for datasets, which requires us to assess the task relevance of each data owner's dataset, where task relevance measures the extent to which the data content is relevant to the task needs.

When the task publisher initiates a federated training task, it sends the federated training model to the quality assessor along with a sample dataset S. The sample dataset represents the data requirement criteria of the task publisher. The sample dataset represents the task publisher's requirement criteria for the data, and therefore the dataset-task relevance quality score of a data owner can be expressed by the similarity index between the local dataset D and the sample dataset S. The relevance of the dataset D and the task M can be expressed as follows:

$$Sim(D,S)=\frac{\sum_{i=1}^{|D|} I(\min_{s_j\in S} dis(d_i,s_j),\delta)}{|D|} \tag{5}$$

$$I(\min_{s_j\in S} dis(d_i,s_j),\delta)=\begin{cases}1,\min_{s_j\in S} sim(d_i,s_j)\le\delta\\ 0,\min_{s_j\in S} sim(d_i,s_j)>\delta\end{cases} \tag{6}$$

In order to train accurate, robust and widely adaptable models during federated learning, the diversity of datasets is crucial. The diversity of a dataset refers to the extent to which the data points in the dataset are distributed and vary in different feature spaces. The diversity of a dataset can be quantified by the average pairwise distance between feature vectors. Specifically, let the dataset D contain n data points, each data point d_i is converted into an m-dimensional feature vector , and the Euclidean distance between feature vectors is denoted as:

$$dis(f(xi),f(xj))=\sqrt{\sum_{k=1}^{m}(f_k(x_i)-f_k(x_j))^2} \tag{7}$$

Define the average pairwise distance as the average of the distances between all data points:

$$Dp(D)=\frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} dis(f(x_i),f(x_j)) \tag{8}$$

where $n(n-1)$ is the total number of data points and the coefficient 2 is due to the fact that the distance matrix is symmetric. The greater the average pairwise distance, the greater the variability and diversity of the data points in the dataset in the feature space.

IV. Arithmetic Resource Metrics

A. Assessment Indicators

In this paper, we argue that arithmetic evaluation for federated learning is a comprehensive concept that needs to include multiple connotations such as computation, storage, network transmission, and arithmetic performance. Among them, computing power is divided into general-purpose computing power represented by CPU and intelligent computing power represented by GPU. Therefore, this paper selects the following five indicators for computing power evaluation: general-purpose computing power, intelligent computing power, storage capacity, transmission capacity, and computing efficiency.

B. Arithmetic Evaluation Model

Users participating in federated learning, whose local devices have certain arithmetic resources, get the score of each arithmetic node by setting up a computational power evaluation model.

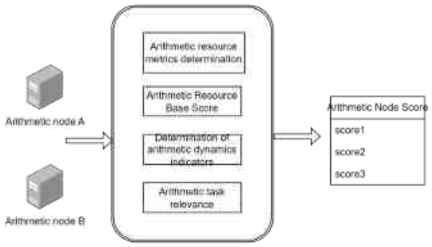
Arithmetic node A
Arithmetic node B
Arithmetic resource metrics determination
Arithmetic Resource Base Score
Determination of arithmetic dynamics indicators
Arithmetic task relevance
Arithmetic Node Score
score1
score2
score3

Figure 4.1Data quality assessment process

The arithmetic metrics are first analysed using the entropy weight method [6] to obtain the arithmetic base score of the arithmetic nodes. However, an arithmetic node may have a high base arithmetic power but perform poorly under dynamic metrics under high load, which may lead to inefficient execution when running certain complex tasks. [7]

Here we choose dynamic metrics such as CPU usage, GPU acceleration ratio, and memory utilisation to measure the actual performance of arithmetic nodes when running specific tasks, so that arithmetic resource judgments are more adaptable to different types of training tasks.

In the process of federated learning, when the task publisher releases the training model task, in addition to specifying the task requirements of the training model, it is also necessary to specify the dynamic arithmetic demand when the task is run s. Assuming that the user task arithmetic demand is $S=\{s_1,s_2,...,s_i\}$, the actual performance of the arithmetic node is $T=\{t_1,t_2,...,t_i\}$, due to the different performance factors of the magnitude of different factors, it is necessary to do the normalisation process, is the result falls within the interval [0,1], and the conversion function is:

$$t^*=\frac{t-\min}{\max-\min} \tag{1}$$

The distance between the two was calculated using the n-dimensional Euclidean distance formula:

$$d(S^*,T^*)=\sqrt{\sum_{i=1}^{n}(s_i^*-t_i^*)^2} \tag{2}$$

When $d(S^*,T^*)$ is smaller, it means that the difference between the two is smaller and more in line with the user's task requirements.

V. Experiment

A. Experimental setup

In this paper, we use the MNIST handwritten digit dataset as the training dataset, and simulate 10 federated training clients to train the federated learning model, and the training model used for federated training is the convolutional neural network (CNN) model for MNIST handwritten digit recognition (Mnist_CNN).

B. Experiments on Data Resource Metrics

In order to explore the impact of data resources on model training accuracy, this paper processes the MNIST dataset by changing the image labels in the MNIST dataset with errors, and setting different scenarios of the dataset mislabelling rate of [0,1], respectively. The accuracy of federated learning model training for datasets with different error rates is shown in Figure 1 below.

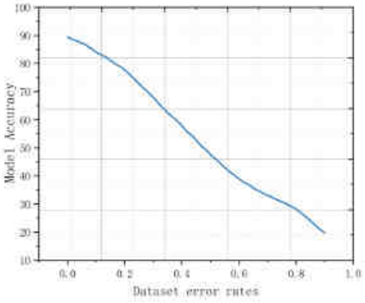

Figure 5.1 Impact of erroneous datasets on model training

Experimental studies have shown that as the error rate of the client dataset increases, its model training accuracy decreases accordingly.

In order to explore the effect of data diversity on the accuracy of federated learning models, we set the client's dataset to have different proportions of homogenised data. The specific dataset is set up in such a way that, while keeping the number of datasets and other quality dimensions consistent, the dataset of the client is set up with label categories from 1 to 10 respectively, and the more categories there are, the lower the degree of homogeneity is. Based on the federal training of datasets with different homogenisation rates, the model accuracy obtained is shown in Figure 2 below.

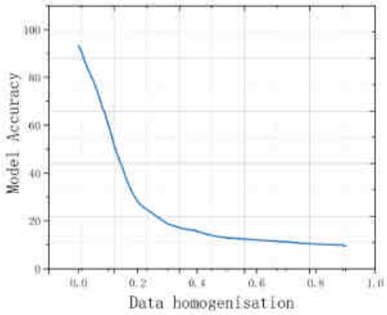

Figure 5.2 Impact of data homogeneity on model accuracy

Based on the training results we can deduce that the higher the degree of data homogeneity and the fewer the categories of images contained in the dataset, the more unfavourable it is for model training in federated learning, and the more the model accuracy is lost.

C. *Experiments on Arithmetic Resource Metrics*

For the arithmetic resource metric we simulate 100 arithmetic nodes, each of which has a base arithmetic metric and a dynamic arithmetic metric, constituting the arithmetic dataset. We calculate the arithmetic score of each client based on its arithmetic metrics. In this paper, we propose the task requester's requirement for arithmetic task relevance, where we use the user's CPU idle rate and storage margin as its dynamic metrics to evaluate its arithmetic task relevance score. The arithmetic requirements of the task requester in federated learning are set as follows: a base performance score of 90 or more, a CPU idle rate of 60%, and a storage margin of 50 GB.

In order to study the level of arithmetic resource utilisation of this scheme, the selection method that only considers the base score of arithmetic resources is compared with the scheme in this paper. As shown in Fig. 3, after considering the arithmetic task relevance score, the utilisation of the arithmetic resources is significantly higher than that of the scheme that only considers the base performance. This situation occurs because the task relevance of both base and dynamic arithmetic is considered, which makes the evaluation of arithmetic resources more accurate, and the matching of federal training tasks with client arithmetic resources more efficient, thus effectively improving the utilisation of arithmetic resources.

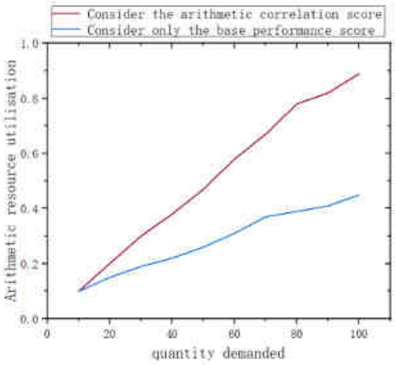

Figure 5.3 Comparison of Arithmetic Resource Utilisation

VI. Conclusion

This chapter looks at the fundamental aspects of incentive design in a federated learning environment, particularly the accurate measurement of participant contribution. Based on an in-depth analysis of the client's role in federated learning and its importance, this chapter provides a comprehensive improvement and refinement of the established contribution measurement methods, addressing the limitations of the previous methods. Starting from the two core dimensions of data resources and arithmetic resources, this chapter meticulously explores a new strategy for measuring the client's contribution and introduces a new concept of task relevance, taking into account the unique attributes of federated learning tasks, with the aim of improving the efficiency of resource allocation. Through experimental design and validation, this chapter verifies the effectiveness of the proposed dual metric mechanism, that high-quality data resources can effectively improve model training accuracy, and through the introduction of the concept of task relevance, the experiments likewise demonstrate that resource utilisation is significantly improved.

References

[1] Yang, Q., Huang, A., Fan, L. et al. Federated Learning with Privacy-preserving and Model IP-right-protection. Mach. Intell. Res. 20, 19–37 (2023). https://doi.org/10.1007/s11633-022-1343-2.

[2] D. Wang, J. Ren, Z. Wang, Y. Wang and Y. Zhang, "PrivAim: A Dual-Privacy Preserving and Quality-Aware Incentive Mechanism for Federated Learning," in IEEE Transactions on Computers, vol. 72, no. 7, pp. 1913-1927, 1 July 2023, doi: 10.1109/TC.2022.3230904.

[3] J. Kang, Z. Xiong, D. Niyato, H. Yu, Y. -C. Liang and D. I. Kim, "Incentive Design for Efficient Federated Learning in Mobile Networks: A Contract Theory Approach," 2019 IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS), Singapore, 2019, pp. 1-5, doi: 10.1109/VTS-APWCS.2019.8851649.

[4] Zhang, J., Li, C., Robles-Kelly, A., and Kankanhalli, M., “Hierarchically Fair Federated Learning”, <i>arXiv e-prints</i>, 2020. doi:10.48550/arXiv.2004.10386.

[5] Li, Yang, Xuewei Chao and Sezai Ercişli. “Disturbed-entropy: A simple data quality assessment approach.” ICT Express 8 (2022): 309-312.

[6] He, Dayi & Xu, Jiaqiang & Chen, Xiaoling. (2016). Information-Theoretic-Entropy Based Weight Aggregation Method in Multiple-Attribute Group Decision-Making. Entropy. 18. 171. 10.3390/e18060171.

[7] T. Song, Y. Tong and S. Wei, "Profit Allocation for Federated Learning," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 2577-2586, doi: 10.1109/BigData47090.2019.9006327.