Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 17, No. 1, 2025

Research on Edge oriented Federated Learning Training Optimization

Wenzhe Zhang, Yong Liu, Xiaoli Song

Henan University of Science and Technology, Luoyang 471000, China

Abstract: Federated learning, as a new distributed computing paradigm, reduces communication bandwidth consumption while ensuring data privacy and security, and can also utilize data from other terminal devices for collaborative training. However, in real edge computing scenarios, the data collected by edge devices usually have certain heterogeneity, which will lead to weight divergence, catastrophic forgetting of knowledge and other phenomena in the model training process of federated learning. Many existing federated learning methods improve from one direction of local client updates and global model updates, inevitably overlooking the impact of the other. Therefore, this article proposes a group based federated continuous learning method (FCL). Firstly, clients with similar data distributions are grouped together to reduce the weight differences between different clients within the same group; Then, during the local training process of the model, intelligent synaptic algorithm terms are introduced, and the learning of each group is modeled as a continuous learning task, in order to integrate knowledge between different local models and improve the model's ability to recognize and analyze old learning tasks. Experiments on the MNIST and CIFAR-10 standard datasets have shown that the model testing accuracy of FCL has improved by approximately 0.31% to 2.17% compared to FedProx, Scaffold, and FedCurve algorithms. Effectively enhancing the anti forgetting ability of the training model, further improving the convergence speed and accuracy of the training model.

Keywords: Federated learning; Continuous learning; Heterogeneous data.

1. Introduction

In recent years, with the rapid development of IoT technology, data at the network edge has been growing rapidly. According to the prediction of well-known consulting firm IDC, the amount of data generated by society will grow at a rate of 50% per year, and in the future, more than half of the data will need to be processed at the edge[1]. This greatly promotes the integration of IoT technology and artificial intelligence[2]. However, in reality, due to bandwidth limitations and data privacy and security issues, the traditional distributed machine learning paradigm has become increasingly impractical. Faced with these issues, federated learning (FL) has received increasing attention[3]. Federated learning technology was first proposed by Google in 2016, with the core idea of achieving multi-party participation in training while ensuring data privacy and security, and solving the problems of data centralization and data silos. In the training process of federated learning, the data of each participant is only processed and used locally, relying on the upload of model parameters to utilize their respective data and collaboratively construct a global model to achieve data sharing without exposing the original data. Federated learning, with its unique distributed training mechanism, not only ensures data privacy and security, but also enhances the model's generalization ability by aggregating diverse data.

However, while federated learning has advantages, it also brings new problems. In the federated learning system, different terminal devices have significant differences in data sources, quantities, and types due to different application scenarios, which leads to data heterogeneity [4]. Moreover, due to the vastly different physical environments and operating scenarios of each device, as well as differences in user collection preferences, the collected data varies greatly in terms of features and distribution and is independent of each other. This type of data with diversity in structure, type, and statistical characteristics is referred to as Non IID data in statistics [5]. This not only makes the model unable to converge quickly, but also greatly reduces the inference accuracy and precision of the global model. Therefore, how to reduce the impact of data heterogeneity on federated learning algorithms and improve the performance of federated learning training models for edge scenarios is of great practical significance for the development and application of federated learning.

In order to solve the problem of decreased model accuracy caused by Non IID data scenarios, relevant scholars have conducted a series of studies. Fraboni et al [6] proposed an aggregate sampling method based on sample size and similarity. This method reduces the weight variance of the training model during clustering by comprehensively considering the sample size and similarity of data features, thereby achieving faster and smoother convergence of the training model. The FedProx proposed by Li et al [7] corrects the deviation in the update direction of the loss function caused by multiple rounds of local training during the model training process, in order to better adapt the model to multivariate data. However, the improvement of FedProx compared to FedAvg is relatively limited. FedCurv [8] adds a penalty term to the loss function during model training, which constrains the local model training process to converge to the shared optimal value. SCAFFOLD [9] introduces knowledge of the global model during local model training to promptly correct local models that deviate from the global training direction and prevent the problem of "customer drift" caused by excessive deviation.

Another research direction is to integrate some technologies into federated learning. Technologies such as meta learning[10], continuous learning[11], and multi task learning [12] are widely used in Non IID federated learning scenarios. One of the good research directions is to apply continuous learning to the field of federated learning.

Shoham[13] and others compared the similarity between federated learning and continuous learning, and applied solutions to deal with catastrophic forgetting in continuous learning to federated learning in non independent and identically distributed data scenarios, effectively solving the problem of local model drift. FedLSD[14] focuses more on updating local model knowledge by using knowledge distillation techniques to extract key knowledge from the global model, in order to enrich the knowledge reserve of the local model. These methods reduce the impact of data heterogeneity, but the distribution of data varies among different clients, and the knowledge learned also differs. There may still be interference from differences during global model aggregation.

Based on the above research results, it is not difficult to find that federated learning optimization methods under non independent and identically distributed data can be improved from two perspectives: global model aggregation and local client updates. However, the vast majority of current solutions only optimize and improve from individual perspectives, which can easily have a negative impact on other aspects when solving a certain problem, ultimately leading to a decline in the overall model quality. We have proposed and implemented FCL to address this phenomenon. FCL divides all participating edge devices into multiple groups based on their data feature maps. The data distribution of participants in a single group is similar, which can avoid the problem of weight divergence and reduced accuracy during random sampling due to different data distributions, and promote faster convergence of the global model training. The data distribution between groups is different. In order to learn the characteristics of different data distributions, we use intelligent synaptic algorithm to fuse the differences between different data. In short, the FCL algorithm reduces the interference of different data distributions during client aggregation through clustering and grouping, and continuously learns beneficial global knowledge through intelligent synaptic algorithms to maximize the convergence speed and model quality of the global model.

2. Group based federated continuous learning method

2.1. Problem description

In federated learning scenarios, the global server and participants are connected through a network, and the training data used for model training is distributed across various edge devices, achieving model training through iterative global aggregation and updates. The optimization objective of federated learning is to minimize the average loss of all samples, as shown in equation (1):

$$f(w) = \sum_{k=1}^{k} \frac{n_k}{n} F_k(w). \quad (1)$$

Among them, $F_k(\boldsymbol{w}) = \frac{1}{\mathrm{n}_k} \sum_{i \in P_k} f_i(\boldsymbol{w})$, w is the neural network parameter, n is the data volume of client K, and N is the total data volume.

$F_k(w)$ represents the distribution information of local data. When all participants are independent and identically distributed, that is, when the data on the client is distributed the same as the overall data, the prediction loss on the client should be expected to be the same as the global prediction loss. After multiple synchronizations, the aggregated global model can approximate the centralized training model; When the data does not meet the assumption of independent and identically distributed, the model fits its own dataset, causing divergence in parameter directions, and as the number of synchronizations increases, the divergence becomes larger, deviating from the global optimal solution during server-side aggregation, as shown in Figure 1.

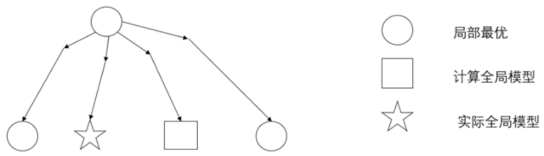

Fig. 1 Model Drift during Non IID Data Training

On the basis of the above theoretical methods, we propose a grouping based federated continuous learning optimization method (FCL), which divides the entire process into two parts: grouping clients based on the similarity of data feature maps, and knowledge fusion for different groups.

2.2. Client grouping

Clients with different data distributions may be affected by the learning of other clients during aggregation. To avoid this issue, we grouped all clients. But we did not train a global model for each group, which resulted in the inability to fully utilize the advantages of large-scale datasets maintained by federated learning. Clients with small data volumes were prone to overfitting, and instead, knowledge fusion was applied to different groups. The specific process of grouping is as follows: Firstly, by utilizing the pre trained model provided in the Torchvision [15] library, data extraction is performed on edge devices and their corresponding feature maps are obtained. Then, the DBSCAN clustering method is used to cluster the collected model parameters, dividing the clients into different groups with similar data distributions in each group. And for FCL, assuming that the distribution of data on the edge device side is fixed, So only one processing is needed. And due to the complexity of the DBSCAN algorithm being O (n), the communication resources consumed by this operation step are much smaller than other operations.

2.3. Continuous learning

How to efficiently integrate knowledge from different groups is the key problem that our algorithm needs to solve after grouping, and continuous learning provides a good idea for solving this problem. The concept and method of continuous learning have opened up a good path to solving this problem. The core idea of continuous learning is to enable the model to retain the knowledge of the previous learning task when learning the next task, and various methods have emerged through development. Intelligent synaptic algorithm, as a highly successful method in the field of continuous learning, stands out in many scenarios due to its unique advantages. This algorithm is based on the principle of regularization and can directly learn from unlabeled data to obtain an importance weight matrix. In addition, the intelligent synaptic algorithm has a small memory footprint, which perfectly fits the resource constrained characteristics of edge computing devices. It can calculate the importance of each parameter to preserve the model's ability to identify and analyze old knowledge, significantly improving the practicality and stability of the model. The loss function of the intelligent synapse algorithm is shown in Equation 2.

$$L(w) = L_n(w) + \frac{\lambda}{2}\sum_i \Omega_{ij}(w_{ij} - w_{ij}^n)^2 \quad (2)$$

Among them, Ln(w) is the loss function of the new task, λ is the hyperparameter with adjustable regularization term, Ωij represents the importance of the parameters, and w_{ij}^n it is the model parameters assigned to the nth training task.

2.4. Local continuous learning

The federated continuous learning method based on intelligent synaptic algorithm proposed in this article. It is necessary to model the stage of model aggregation as a continuous learning task and ensure the orderliness of the continuous learning task. In the FCL algorithm, all grouping orders are required to participate in training for each communication round, and clients are randomly selected from the same group for weighted aggregation. When selecting groups for training for the first time, there is no need to add the memory aware synaptic algorithm term, so the loss function is the cross entropy loss function, as shown in Equation 3.

$$L(w) = -\frac{1}{n}\sum_x (y\ln a + (1-y)\ln(1-a)) \quad (3)$$

Among them, n is the sample size, y is the label vector, and a is the output vector. Except for the initial selection of groups for training, the model parameters received by the local model during local updates are all from the previous group, which reduces the current training model's ability to recognize and analyze old knowledge. To solve this problem, this paper introduces the intelligent synapse algorithm into the local training stage of the local model, which can reduce the degree of change in important parameters of old knowledge. The data distribution in our same group is similar, so we can directly randomly select clients to update Ωij. The specific process is shown in Table 1.

Tab.1 FCL algorithm training steps

FCL Algorithm steps
Input: Local training round E, communication round C, initialization of global model w^0, proportion of participating clients r, coefficient γ, device grouping G
Output: Global Model w^t
Server side execution
1.While {C-1 > 0} do
2. For each $G_i \in G$ do
3. nums ← max(r·G_i,1)
4. S_t←random subset of nums clients in G_i
5. For each clint $k \in S_t$ in parallel do
6. w_{t+1}← ClientUpdate(k,w_t ,Ω)
7. $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$
8. IF non first-time selection of clients within the group then
9. S_t←random subset of 1 clients in G_i
10. Ω←calculate（w_{t+1}）
Client Execution
11.ClientUpdate(k,w_t ,Ω)
12. β← (split X_k into batches of size B)
13. For each local epoch i from 1 to E do
14. For batch $b \in \beta$ do
15. L_k（w）$\leftarrow \ell(w;b)$
16. IF Ω exist then
17. L_k（w）←L_k（w）+γ$\sum_i \Omega_i(w_i - w_{t,i})$
18. argmin$_w L_k$（w）
19.return w to server

Overall, the FCL method proposed in this article reduces the impact of weight divergence within groups by grouping clients with similar data distributions into the same group; Introducing intelligent synaptic algorithms during the local training stage of the model to integrate knowledge between different local models and improve the model's ability to resist forgetting old learned knowledge, thereby enhancing the performance and accuracy of the model.

3. Experimental results and analysis

3.1. Experimental setup

To verify the performance of the FCL algorithm proposed in this article, experiments were conducted using the MNIST and CIFAR-10 datasets in this section. At the same time, in order to reflect the characteristics and performance of FCL algorithm in edge computing scenarios with heterogeneous data, we compare it with some federated learning algorithms. Among them, the FedProx algorithm introduces a proximal term to limit the difference between local and global models. The Scaffold algorithm introduces a control variable to alleviate the impact of non independent and identically distributed data on the training model. FedCurve is similar to FedProx, except that the regularization term is changed to the regularization term of the EWC algorithm. By continuously learning the regularization term, it overcomes the catastrophic forgetting of federated learning under data heterogeneity. FedCurve performs better when λ=2.0 in the original text, so we directly set it to λ=2.0. For the hyperparameters of the FCL algorithm, if set too small, it has no effect on local updates. If set too large, the model updates slowly. Referring to the setting method of FedFMC[16], set it to λ=1/number of clients. At the same time, in order to better conform to the real edge computing scenario, this chapter simulates the non IID distribution of skewed labels in different client data sets through the Dirichlet distribution in reference[17]. We directly set the parameter α to 0.3 and 0.6 to complete the sampling of the dataset, and then use this dataset to evaluate the performance of FCL in training models with different degrees of non independent and identically distributed data.

The experimental parameters are set as follows: in local training, the SGD parameter in the optimizer used is set to a learning rate of 0.01, with 5 local training epochs. In the comparative experiment, the MNIST dataset has 50 communication epochs, the CIFAR-10 dataset has 100 communication epochs, the total number of clients is 200, the sample rate for training is set to 0.1, the sample input dimension is 64, and the batch size is set to 10.

3.2. Ablation experiment

The FCL algorithm proposed in this article is mainly divided into two parts: (1) grouping based on the similarity of data features, and (2) introducing the intelligent synapse algorithm into the local update stage of the training model. To verify the effectiveness of this method, ablation experiments were conducted on both parts of the method. This article uses the MNIST dataset for experiments, with communication rounds set to 200 and other experimental parameters unchanged. The experiments are conducted using FedAvg algorithm (FedAvg), FedAvg algorithm with intelligent synapse algorithm (FedAvgMas), FedAvg algorithm with client grouping strategy (GFedAvg), and FedAvg algorithm with client grouping strategy and intelligent synapse algorithm (GFedAvgMas). The results are shown in Figure 2.

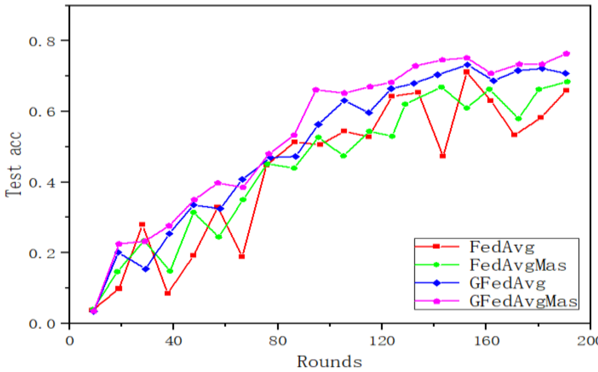

Fig.2 The effectiveness of grouping and local updates

(1) The effectiveness of grouping strategies. The grouping strategy is mainly aimed at reducing the impact of different data distributions on clients during model aggregation. To verify the effectiveness of this method, FedAvg and FedAvgMas algorithms randomly select clients, while GFedAvg and GFedAvgMas constrain the subsequent client selection through client grouping. From the figure, we can see that the algorithm using client grouping strategy improves the accuracy of the model by about 6.26% to 13.37% compared to the algorithm randomly selecting clients; Although there were no significant differences between the various methods in the first half of the communication rounds, the algorithm using client grouping strategy was more stable during the improvement process; After 100 rounds of communication, it can be seen that the model accuracy of the algorithm using grouping strategy is always higher than other methods; And the method of randomly selecting clients has always had significant fluctuations. This indicates that grouping strategy can effectively reduce the mutual interference between clients with different data distributions during aggregation, thereby improving the accuracy of the training model.

(2) The effectiveness of intelligent synaptic algorithms. During the local update stage of the training model, FedAvgMas and GFedAvgMas, which added intelligent synaptic algorithm terms, improved the testing accuracy of the training model by 6.21% and 7.43% respectively compared to FedAvg and GFedAvg. This indicates that adding intelligent synaptic algorithm during the local update stage of the model can effectively improve the anti forgetting ability of the local model, thereby alleviating the catastrophic forgetting problem that occurs during model training.

3.3. Comparative experimental results

The accuracy comparison between FCL and other methods on datasets with different degrees of non independent and identically distributed data is shown in Table 2. It is evident that the FedCurve algorithm and FCL algorithm, which combine continuous learning algorithms, can obtain a high-performance global model faster than the other two methods; When faced with Non IID data, the FCL algorithm can achieve higher accuracy compared to other algorithms, indicating that the method proposed in this paper can effectively reduce the impact of non independent and identically distributed data on the training model; When the data distribution is closer to independent and identically distributed data, the FCL algorithm does not have significant advantages compared to other algorithms. However, even when facing independent and identically distributed data, the FCL algorithm can improve the performance of the model due to its anti forgetting ability. Therefore, FCL has more significant advantages than other federated learning in edge computing scenarios with heterogeneous data.

Tab. 2 Comparison of accuracy under different levels of Non IID

	MNIST		CIFAR-10	
	α=0.3	α=0.6	α=0.3	α=0.6
FedProx	88.83%	90.34%	42.42%	44.73%
Scafflod	88.79%	88.63%	42.50%	44.82%
FedCurv	89.44%	90.94%	43.27%	45.53%
FCL	90.64%	91.12%	43.84%	45.37%

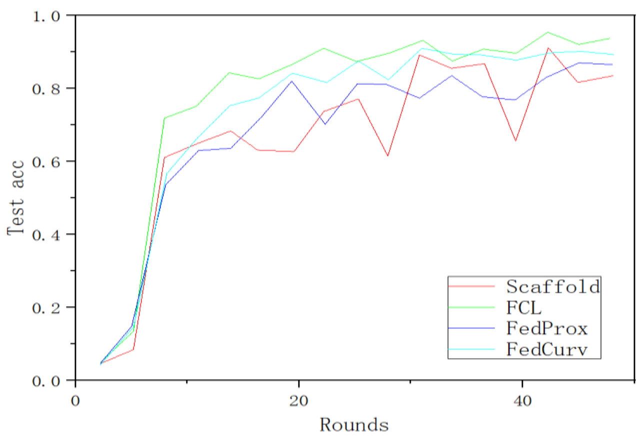

Fig.3 Comparison of accuracy of different algorithms on the MNIST dataset

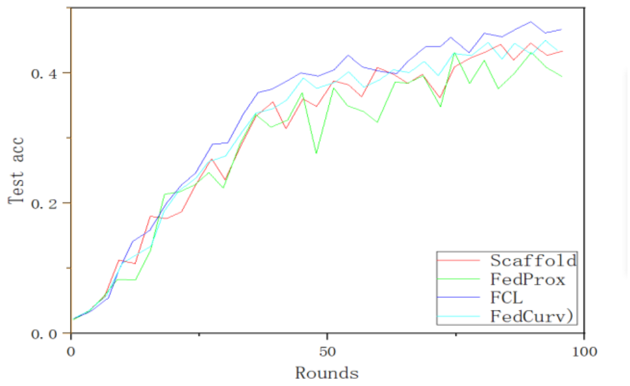

Fig.4 Comparison of accuracy of different algorithms on the CIFAR-10 dataset

The changes in model testing accuracy of various algorithms on the MNIST dataset and CIFAR-10 with α=0.3 are shown in Figures 3 and 4. It can be seen that the FCL method proposed in this paper not only achieves faster model convergence on both the MNIST dataset and CIFAR-10 dataset, but also outperforms other methods in terms of final model accuracy. On the CIFAR-10 dataset, the accuracy of the FCL algorithm was improved by 0.31%, 0.54%, and 0.82% compared to other algorithms, demonstrating the effectiveness of the proposed method. The FCL algorithm introduces intelligent synaptic algorithm during local updates to limit the degree of change in important parameters. Therefore, compared with other algorithms, it not only has more stable changes, but also has higher model accuracy.

4. Conclusion

In the edge computing environment with heterogeneous data, different local models usually affect each other in the training process, which will not only reduce the edge device's ability to identify and analyze old data, but also cause local model drift, thus affecting the performance of the global model. In response to this issue, this article innovatively proposes a federated continuous learning method based on intelligent synaptic algorithm. This method divides clients with similar data distributions into the same group to reduce the impact of different weight differences within the group; Afterwards, during the local training stage of the model, intelligent synaptic algorithms are introduced to model the model training problem as a continuous learning task. To integrate knowledge between different local models, reduce the magnitude of changes in parameters that are more important for training models on old tasks, improve the anti forgetting ability of the training model, and thus make the global model approach the global optimal solution. Finally, experimental results have shown that this method can reduce the impact of Non IID data on local model training, alleviate the catastrophic forgetting problem of the training model, and effectively improve the training accuracy of the global model.

The method architecture in this article takes into account the heterogeneity of client data in real-world scenarios, therefore it has a certain universality and can be applied in scenarios where multiple clients train together. In future work, we will also focus on how to improve the anti forgetting ability of the training model while enhancing its plasticity when facing new learning tasks, in order to design better performing federated learning algorithms.

Acknowledgements

Key Research and Development and Promotion of Special (Science and Technology) Project of Henan Province, China (No. 232102210153)

References

[1] Wang M , Deng W .Deep face recognition: A survey [J].Neurocomputing, 2022.

[2] Lim W Y B , Luong N C , Hoang D T ,et al.Federated Learning in Mobile Edge Networks: A Comprehensive Survey[J].IEEE Communications Surveys & Tutorials, 2020, PP(99):1-1.

[3] Ramaswamy S , Mathews R , Rao K ,et al.Federated Learning for Emoji Prediction in a Mobile Keyboard[J].arXiv, 2023.

[4] Wang H , Kaplan Z , Niu D ,et al.Optimizing Federated Learning on Non-IID Data with Reinforcement Learning[C]//IEEE INFOCOM 2021 - IEEE Conference on Computer Communications.IEEE, 2021.

[5] Zhao Y , Li M , Lai L ,et al.Federated Learning with Non-IID Data[J].2018.

[6] Fraboni Y , Vidal R , Kameni L ,et al.Clustered Sampling: Low-Variance and Improved Representativity for Clients Selection in Federated Learning[J].2023.

[7] Li T , Sahu A K , Zaheer M ,et al.Federated Optimization in Heterogeneous Networks[J]. Proceedings of Machine Learning and Systems, 2021, 2: 429-450.

[8] Shoham N , Avidor T , Keren A ,et al.Overcoming Forgetting in Federated Learning on Non-IID Data[J]. 2021.

[9] KARIMIREDDY S P, KALE S, MOHRI M, et al. Scaffold: Stochastic controlled averaging for on-device federated learning. [J]. 2020.

[10] Jiang Y , Konen J , Rush K ,et al.Improving Federated Learning Personalization via Model Agnostic Meta Learning[J].arXiv, 2021.

[11] Hu W, Qin Q, Wang M, et al. Continual learning by using information of each class holistically[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2023.

[12] CORINZIA L, BEURET A, BUHMANN J M. Variational federated multi-task learning. [J]. arXiv preprint arXiv:1906.06268, 2020.

[13] Shoham N, Avidor T, Keren A, et al. Overcoming forgetting in federated learning on non-iid data [EB/OL].[2021-10-17].

[14] Lee G,Shin Y,Jeong M,et al.Preservation of the global knowledge by not-true self knowledge distillation in federatedlearning [EB/OL].[2023-11-29]. https://arxiv.org/abs/ 2106.03097v2.

[15] TORCHVISIONMODEL. Pre-trained models from torchvision. [EB/OL]. 2023.

[16] Kopparapu K, Lin E. FedFMC:Sequential Efficient Federated Learning on Non-iid Data[EB/OL].[2023-6-19].

[17] HSU T M H, QI H, BROWN M. Measuring the effects of non-identical data dis-tribution for federated visual classification. [J]. arXiv preprint arXiv:1909.06335,2021.