Fuzzy Clustered Federated Learning Under Mixed Data Distributions

Extended Abstract

Peng Tang
Shanghai Jiao Tong University
Shanghai, China
tangpeng@sjtu.edu.cn

Lifan Wang
Shanghai Jiao Tong University
Shanghai, China
intefirm@sjtu.edu.cn

Weidong Qiu
Shanghai Jiao Tong University
Shanghai, China
qiuwd@sjtu.edu.cn

Zheng Huang
Shanghai Jiao Tong University
Shanghai, China
huang-zheng@sjtu.edu.cn

Qiangmin Wang
Shanghai Jiao Tong University
Shanghai, China
qmwang@sjtu.edu.cn

ABSTRACT

Federated learning (FL) is deeply troubled by non-independent and identically distributed (non-IID) data, leading to suboptimal training results. Clustered FL partitions clients' unique data into different clusters to reduce the heterogeneity among clients. Current approaches are unable to eliminate the impact of data heterogeneity and provide personalized models to client devices. By assuming the clients' data can be divided into different data distributions, we propose a novel fuzzy clustered FL method. We partition the client's data and generate a personalized model for each client. The experiments demonstrated that our method achieved excellent results. In the case of N clusters, our method achieved a communication cost reduction of $1/N$ compared to the SOTA methods, while improving performance by 10.4% on CIFAR-10.

KEYWORDS

Federated Learning

ACM Reference Format:
Peng Tang, Lifan Wang, Weidong Qiu, Zheng Huang, and Qiangmin Wang. 2024. Fuzzy Clustered Federated Learning Under Mixed Data Distributions: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

FL [2] allows multiple parties to collaboratively train models without sharing private data, thus avoiding the potential harm to privacy caused by data sharing. Due to the highly decentralized system architecture, data between different clients may be non-IID, such as different data sizes and class distributions. This poses significant challenges to FL, such as overfitting or bias issues.

Clustered FL [3, 7] assumes that different client data may share same data distribution, which means that clients can be divided into different clusters. Each cluster obtains a cluster model with better performance on it's data distribution. By employing hierarchical partitioning, clustered FL groups clients with the same data distribution into the same cluster.

But real-world clients are unlikely to have exactly the same data distribution. More commonly, client's data comes from multiple distributions [4], so each client's data distribution can be regard as a mixture of multiple distinguishable distributions. This setting allows us to perform a more fine-grained analysis of the data. We propose a fuzzy clustered FL method, where client's data is split across multiple clusters. By using this fuzzy cluster approach, we can more accurately aggregate similar data together, rather than directly classifying all user data into one cluster, and each client will receive a personalized local model related to its data distribution. Through iterative clustering and model updates, knowledge sharing is indirectly achieved between different cluster distributions.

By relaxing the assumption of non-IID, we propose a fuzzy clustered FL method which is more applicable to real-world scenarios. It introduces less overhead and has higher communication efficiency. In multiple non-IID datasets, our method can identify natural clusters and accurately partition clients' data into different clusters, effectively reducing the impact of data heterogeneity. Without hyperparameter tuning, Our method customizes for specific distributions within clusters, achieving better accuracy and generalization capabilities, and generates personalized local models for clients.

2 FUZZY CLUSTERED FEDERATED LEARNING

Due to the non-shared nature of data in FL, the cluster membership is an unknown parameter. We introduce contrastive learning [1] for clustering. We use a contrastive projection head (CPH) $g(\cdot)$. The output z_i^a is the soft labels belonging to each cluster, the value of the i-th dimension can be interpreted as the probability that the vector belongs to the i-th cluster. Then a positive pair generated by data augmentation is used to calculate the cluster contrastive loss ℓ_{clu} for unsupervised learning. To decrease the cosine distance between (z_i^a) and (z_i^b), we define cluster contrastive loss as follows, where τ denotes a temperature parameter:

$$\ell_i^a = -log \frac{exp(s(z_i^a, z_i^b))/\tau}{\sum_{j=1}^{N} [exp(s(z_i^a, z_j^b))/\tau + exp(s(z_j^a, z_j^b))/\tau]} \quad (1)$$

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Dataset	CIFAR-10 (7:3)				CIFAR-10 (9:1)				CIFAR-100				EMNIST			
Metrics	Cluster1	Cluster2	Client	MSE	Cluster1	Cluster2	Client	MSE	Cluster1	Cluster2	Client	MSE	Cluster1	Cluster2	Client	MSE
FedAvg	0.866	0.671	0.709	-	0.865	0.660	0.724	-	0.467	0.529	0.503	-	0.792	0.703	0.808	-
FedSoft	0.849	0.656	0.683	0.181	0.825	0.612	0.690	0.298	0.443	0.467	0.466	0.580	0.767	0.676	0.813	0.580
FedEM	0.656	0.481	0.576	0.072	0.650	0.403	0.547	0.300	0.267	0.429	0.432	0.171	0.320	0.281	0.192	0.154
Our	**0.896**	**0.740**	**0.754**	**0.013**	**0.891**	**0.720**	**0.760**	**0.190**	**0.500**	**0.633**	**0.535**	**0.150**	**0.820**	**0.759**	**0.829**	**0.039**

Table 1: Average client accuracy, cluster model's accuracy and MSE of membership.

By minimizing the cluster contrastive loss, the model will have clustering capability. CPH and the original classification network share the same backbone and simultaneously optimized. The local objective function ℓ consists the cluster contrastive loss ℓ_{clu} and the cross-entropy loss ℓ_{sup}:

$$\ell = \ell_{sup} + \beta\ell_{clu} \tag{2}$$

Considering there are N cluster models $\{C_i\}$ and M clients, our algorithm makes two major improvements. Firstly, client k uses local data to update the local model, and uses CPH to determine the number of data points $\{n_i^k\}$ belonging to cluster i as the cluster membership. After local training, the client uploads the model, then the server performs model aggregation for each cluster model:

$$C_i = \sum_{k=1}^{M} \frac{n_i^k}{\sum_{j=1}^{M} n_i^j} w_k \tag{3}$$

Secondly, a client should receive more weight in the model aggregation if they have more data points belonging to a certain cluster. By using the cluster membership for each client as the weight for weighted averaging, we can partition the data belonging to this cluster into a single cluster and reduce the impact of heterogeneous data. After the server completes the aggregation of the cluster model, a model is generated based on the client's local data distribution via:

$$w_k = \sum_{i=1}^{N} \frac{n_i^k}{\sum_{j=1}^{N} n_j^k} C_i \tag{4}$$

Using global mixed updates and model aggregation, we ensure knowledge exchange among different clusters while minimizing the impact of data heterogeneity on the model.

3 EXPERIMENTS

Our approach is built upon a discussion of the specific form of non-IID data. In clustered FL, data is assumed to have certain partition relationships. We believe that a more fitting scenario, taking the CIFAR-10 dataset as an example, is that some clients have more pictures of transportation vehicles while others have more pictures of animals. Through this natural partition relationship, we can learn cluster allocation in an unsupervised manner. Our algorithm is evaluated on three public and wide-used datasets: CIFAR-10, EMNIST, and CIFAR-100. We use FedAvg [5], FedSoft [6], and FedEM [4] as baselines. For clustered FL, we compare the mean squared error between the cluster membership probabilities and the actual data distribution. As shown in Table 1, our method achieves better results than FedAvg and other clustered FL methods. The results show that our method achieves higher accuracy both globally and locally. At the same time, compared with the existing clustered FL methods, our clustering results are more accurate.

Two ablation studies are carried out to further understand the effect of cluster contrastive loss and the effect of fuzzy clustered FL. Firstly, we added cluster contrastive loss ℓ_{clu} to the local loss function of FedAvg. Secondly, we assumed that FedAvg would also generate two cluster models, and the cluster membership is known by the server. This enables us to use the same clustered FL approach on FedAvg. By aggregating the models based and generating local personalized models on the membership, we can perform better clustered FL on FedAvg.

Dataset	CIFAR-10 (7:3)		
Metrics	Cluster1	Cluster2	Client
FedAvg	0.866	0.671	0.709
FedAvg $+\ell_{clu}$	0.871	0.699	0.724
FedAvg $+clusteredFL$	0.879	0.698	0.704
our	**0.896**	**0.740**	**0.754**

Table 2: Results of ablation experiments

Table 2 shows the result of two ablation studies. Firstly, by utilizing cluster contrastive loss for representation learning, the model gains enhanced feature extraction capabilities, leading to improved classification performance. The second study generate two cluster models correspond to different categories of data, which the first model performs better on images of transportation vehicles, and the second performs better on animals. Using cluster membership as weights to perform global aggregation and local update, even on FedAvg, we can obtain two more professional cluster model.

4 CONCLUSION

We proposed a new method for fuzzy clustered federated learning. In a more realistic non-IID environment, our method effectively overcomes the heterogeneity of data. It allows participating users to train personalized models while achieving excellent cluster models on a specific category of data through global training. Compared to existing clustering federated learning methods and other approaches to address heterogeneous data, our method outperforms in both global and client-side performance. Additionally, under the premise of not sharing federated learning data, our method only slightly increases the local workload while almost not changing the communication volume. Our method only adds a lightweight clustering module, which can easily integrate with existing federated learning frameworks and provides good scalability.

REFERENCES

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*. PMLR, 1597–1607.

[2] Jakub Konečný, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. 2016. Federated learning: Strategies for improving communication efficiency. arXiv:1610.05492 [cs.CL]

[3] Yishay Mansour, Mehryar Mohri, Jae Ro, and Ananda Theertha Suresh. 2020. Three approaches for personalization with applications to federated learning. *arXiv preprint arXiv:2002.10619* (2020).

[4] Othmane Marfoq, Giovanni Neglia, Aurélien Bellet, Laetitia Kameni, and Richard Vidal. 2021. Federated multi-task learning under a mixture of distributions. *Advances in Neural Information Processing Systems* 34 (2021), 15434–15447.

[5] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.

[6] Yichen Ruan and Carlee Joe-Wong. 2022. Fedsoft: Soft clustered federated learning with proximal local updating. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 8124–8131.

[7] Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. 2020. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE transactions on neural networks and learning systems* 32, 8 (2020), 3710–3722.