

FedClust: Optimizing Federated Learning on Non-IID Data through Weight-Driven Client Clustering

Md Sirajul Islam¹, Simin Javaherian¹, Fei Xu², Xu Yuan³, Li Chen¹, and Nian-Feng Tzeng¹

¹School of Computing and Informatics, University of Louisiana at Lafayette, USA

²School of Computer Science and Technology, East China Normal University, China

³Department of Computer and Information Sciences, University of Delaware, USA

Emails: {md-sirajul.islam1, simin.javaherian1, li.chen, tzeng}@louisiana.edu, fxu@cs.ecnu.edu.cn, xyuan@udel.edu

Abstract—Federated learning (FL) is an emerging distributed machine learning paradigm enabling collaborative model training on decentralized devices without exposing their local data. A key challenge in FL is the uneven data distribution across client devices, violating the well-known assumption of independent-and-identically-distributed (IID) training samples in conventional machine learning. Clustered federated learning (CFL) addresses this challenge by grouping clients based on the similarity of their data distributions. However, existing CFL approaches require a large number of communication rounds for stable cluster formation and rely on a predefined number of clusters, thus limiting their flexibility and adaptability. This paper proposes *FedClust*, a novel CFL approach leveraging correlations between local model weights and client data distributions. *FedClust* groups clients into clusters in a one-shot manner using strategically selected partial model weights and dynamically accommodates newcomers in real-time. Experimental results demonstrate *FedClust* outperforms baseline approaches in terms of accuracy and communication costs.

Index Terms—Federated Learning, Clustered Federated Learning, Personalization, Non-IID Data, Client Clustering

I. INTRODUCTION

Federated learning (FL) has become a promising solution for analyzing and processing the massive data generated by various edge devices. Traditional machine learning approaches fall short in handling this exponential increase in data, as they require transmitting large volumes of user data to centralized cloud servers. This incurs prohibitive communication costs and raises privacy concerns. FL enables a set of participating devices to collaboratively train a globally shared model under the coordination of a central server, without exposing their local data. Due to its superior privacy-preservation implications, FL has been widely adopted in diverse areas.

However, deploying FL often involves a significant number of devices that generate heterogeneous data due to varying usage patterns of users. The presence of heterogeneous data across client devices breaks the conventional assumption of independent-and-identically-distributed (IID) training data, raising the new challenge of non-IID data distribution in the FL paradigm. Such a data heterogeneity issue not only increases the overall communication cost but also degrades global model performance [1], [2], increasingly drawing research attention

to mitigate the adverse impact of non-IID data on FL [3]–[5]. Several FL approaches, including personalization [6]–[8], clustering [3], [4], [9], and device selection [10]–[12] have been proposed in the literature to alleviate the impact of non-IID data. Despite the significant improvements CFL-based approaches have shown over FedAvg [13] in handling non-IID data, they still lack efficiency due to limitations in clustering strategies. We identify key limitations of existing CFL methods, including the difficulty in pre-determining cluster counts [3], [4], the need for larger communication rounds to form stable clusters [14], and the necessity of utilizing all model weights for clustering [3], [14].

To address the aforementioned limitations, we propose a novel clustered federated learning method, named *FedClust*, which leverages our insight into the implicit relationship between the local model weights and the underlying data distribution on a client device.

II. BACKGROUND AND MOTIVATION

Federated learning (FL) is a privacy-preserving framework enabling distributed clients to collaboratively train a shared global model without sharing their local data. In FL, clients receive initial global model parameters from a central server, train it for a few local iterations using their local data, and send model updates back to the server for aggregation. After aggregation, the server sends it to clients for further training. The above process continues until achieving a certain level of model accuracy or a pre-specified number of communication rounds. The server has no prior knowledge about the data distribution across devices as it cannot access the raw data stored in clients. In particular, the goal of *FedAvg* is typically to minimize the following objective function:

$$\min_{\theta} F(\theta) \triangleq \sum_{i=1}^m \frac{n_i}{N} F_i(\theta) \quad (1)$$

Here, m is the set of participating clients and client i has local dataset $\mathcal{D}_i$, where $n_i = |\mathcal{D}_i|$ and $N = \sum_{i=1}^m n_i$. The local objective functions of clients can be defined as the empirical loss over their local data $\mathcal{D}_i$, i.e., $F_i(\theta) = \frac{1}{n_i} \sum_{j=1}^{n_i} f_{j_i}(\theta; x_{j_i}, y_{j_i})$, where n_i is the number of each client local samples. We

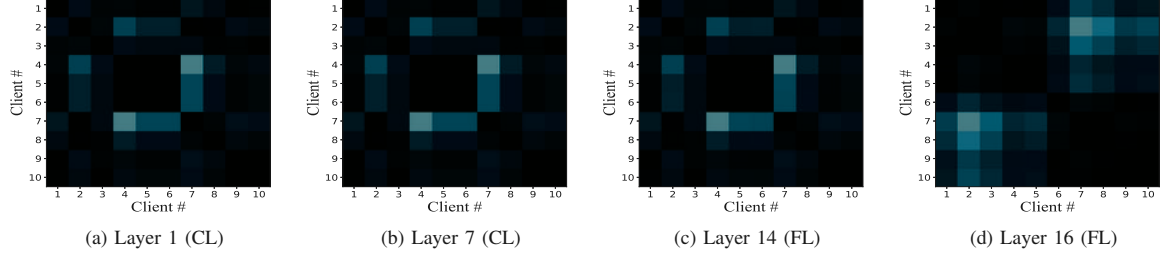


Fig. 1: Illustration of the distance matrices calculated using different layer weights (CL: convolutional layer, FL: fully connected layer). Lighter color indicates smaller distances, i.e., the two models are more similar.

experimentally investigate the impact of model weights from various layers on the underlying data distribution in FL by performing a multi-class image classification task on the CIFAR-10 [15] dataset with VGG16 [16]. To simulate non-IID data among clients, we assume 10 different clients and group them into two clusters based on their local label categories, e.g., $\mathcal{G}_1 = \{1, 2, \dots, 5\}$ and $\mathcal{G}_2 = \{6, 7, \dots, 10\}$. From Fig. 1, we observe that the final layer weights implicitly represent the underlying data distribution of clients. The clustering structure of the clients is clearly observed in Fig. 1(d). Based on the above experiment and previous studies [17], [18], we can conclude that the final layer or the layer with the classifier function reflects the model difference caused by non-IID data. In addition, clients with similar data distributions tend to train the model in a similar manner, resulting in closer distances among final layer weights.

III. THE FRAMEWORK OF FEDCLUST

We present the framework of the proposed FedClust as follows. As shown in Fig. 2, the framework of FedClust can be described as follows.

- First, the server broadcasts the initial global model parameters to all available clients.
- The participating clients locally train the model based on their data, and upload the updated final layer weights to the server to represent their underlying data distribution.

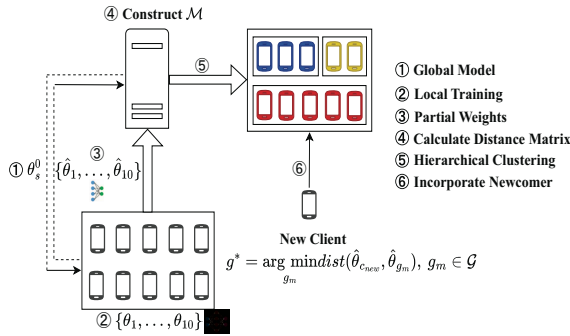


Fig. 2: An overview of FedClust.

TABLE I: Test accuracy comparisons of different approaches over different datasets for Non-IID Dir (0.1).

Method	CIFAR-10	FMNIST	SVHN
FedAvg	38.25 $\pm$ 2.98	81.93 $\pm$ 0.64	61.26 $\pm$ 0.95
FedProx	51.60 $\pm$ 1.40	74.53 $\pm$ 2.16	79.64 $\pm$ 0.80
CFL	41.50 $\pm$ 0.35	74.01 $\pm$ 1.19	61.96 $\pm$ 1.58
IFCA	50.51 $\pm$ 0.61	84.57 $\pm$ 0.41	74.57 $\pm$ 0.40
PACFL	51.02 $\pm$ 0.24	85.30 $\pm$ 0.28	76.35 $\pm$ 0.46
FedClust	60.25 $\pm$ 0.58	95.51 $\pm$ 0.17	78.23 $\pm$ 0.30

- The server then computes the proximity matrix between models based on the Euclidean distance among final layer weights uploaded by each client.
- Finally, the server employs agglomerative hierarchical clustering (HC) [19] on the proximity matrix $\mathcal{M}$ to group clients with similar data distribution into the same cluster.
- The above clustering process is done in one communication round. From the next round, the workflow of FedClust is similar to FedAvg [13].

IV. PERFORMANCE EVALUATION

To evaluate the performance of FedClust, we consider LeNet-5 [20] model on CIFAR-10 [15], Fashion MNIST (FMNIST) [21], and SVHN [22] datasets. We simulate non-IID scenarios using Non-IID Dir (0.1) data heterogeneity settings as in [23]. We consider following baseline approaches: FedAvg [13], FedProx [24], IFCA [3], PACFL [9] and CFL [14]. Results in Table I demonstrate the effectiveness of FedClust across various data heterogeneity scenarios with different models and datasets has been left as part of our future work.

V. ACKNOWLEDGEMENT

The research is supported in part by the NSF under grants OIA-2019511, OIA-2327452, 2348452, and 2315613, in part by the Louisiana BoR under LEQSF(2019-22)-RD-A-21, in part by the NSFC under 62372184, and the Sci. and Tech. Commission of Shanghai Municipality under 22DZ2229004.

REFERENCES

- [1] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-iid data," *arXiv preprint arXiv:1907.02189*, 2019.
- [2] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," *arXiv preprint arXiv:1806.00582*, 2018.
- [3] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, "An efficient framework for clustered federated learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19 586–19 597, 2020.
- [4] X. Ouyang, Z. Xie, J. Zhou, J. Huang, and G. Xing, "Clusterfl: a similarity-aware federated learning system for human activity recognition," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021, pp. 54–66.
- [5] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.
- [6] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," *Advances in Neural Information Processing Systems*, vol. 33, pp. 3557–3568, 2020.
- [7] T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: Fair and robust federated learning through personalization," in *International Conference on Machine Learning*. PMLR, 2021, pp. 6357–6368.
- [8] J. Luo, M. Mendieta, C. Chen, and S. Wu, "Pgfd: Personalize each client's global objective for federated learning," *International Conference on Computer Vision*, 2023.
- [9] S. Vahidian, M. Morafah, W. Wang, V. Kungurtsev, C. Chen, M. Shah, and B. Lin, "Efficient distribution similarity identification in clustered federated learning via principal angles between client data subspaces," in *Proceedings of the AAAI conference on artificial intelligence*, 2022.
- [10] Y. J. Cho, J. Wang, and G. Joshi, "Client selection in federated learning: Convergence analysis and power-of-choice selection strategies," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.
- [11] F. Lai, X. Zhu, H. V. Madhyastha, and M. Chowdhury, "Oort: Efficient federated learning via guided participant selection," in *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2021, pp. 19–35.
- [12] S. Javaherian, S. Panta, S. Williams, M. S. Islam, and L. Chen, "Fedfair³: Unlocking threefold fairness in federated learning," in *IEEE International Conference on Communications (ICC)*, pp. 1–7, 2024.
- [13] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.
- [14] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 8, pp. 3710–3722, 2020.
- [15] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [17] M. Long, Y. Cao, Z. Cao, J. Wang, and M. I. Jordan, "Transferable representation learning with deep adaptation networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 12, pp. 3071–3085, 2018.
- [18] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in neural information processing systems*, vol. 27, 2014.
- [19] W. H. Day and H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *Journal of classification*, vol. 1, no. 1, pp. 7–24, 1984.
- [20] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [21] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.
- [22] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," 2011.
- [23] Q. Li, Y. Diao, Q. Chen, and B. He, "Federated learning on non-iid data silos: An experimental study," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 2022, pp. 965–978.
- [24] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.