Bold but Cautious: Unlocking the Potential of Personalized Federated Learning through Cautiously Aggressive Collaboration (Supplemental Material)

Xinghao Wu[1], Xuefeng Liu[1,2], Jianwei Niu *[1,2,3], Guogang Zhu[1], and Shaojie Tang[4]

[1]State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China

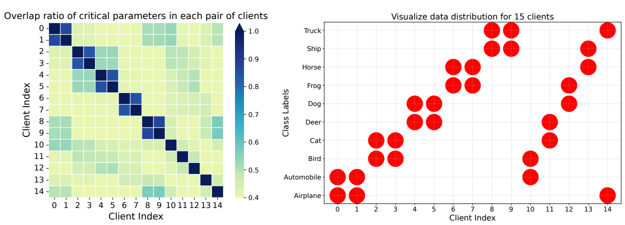
{wuxinghao, liu_xuefeng, niujianwei, buaa_zgg}@buaa.edu.cn

[2]Zhongguancun Laboratory, Beijing, China

[3]Zhengzhou University Research Institute of Industrial Technology, School of Information Engineering, Zhengzhou University, Zhengzhou, China

[4]Jindal School of Management, University of Texas at Dallas, USA

shaojie.tang@utdallas.edu



(a) (b)

Figure 1. An experiment to illustrate that clients with similar data distribution have similar locations of critical parameters.

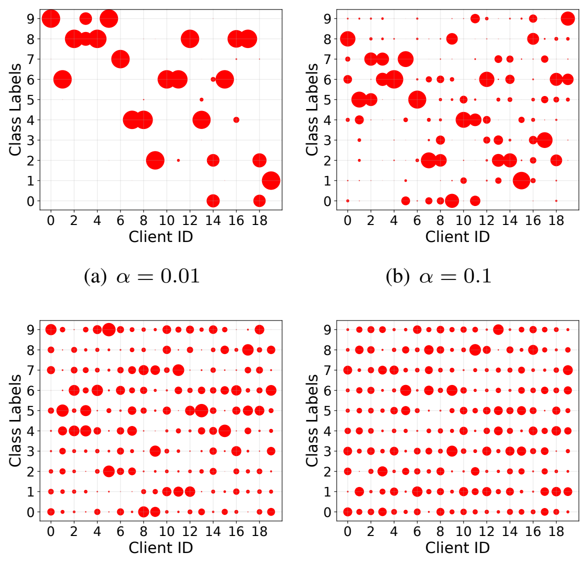
A. Relationship Between Client Data Distribution Similarity and Location Overlap Ratio of Critical Parameter

As we describe in section 1, a key factor affecting client parameter collaboration is the data distribution difference between clients (i.e., Ψ in Eq.(1)). However, due to privacy constraints in FL, we can not know the data distribution of clients, which brings a challenge when implementing Eq.(1). In Figure 1(b) of section 1, we find that the sensitivity of parameters is related to the data distribution. Therefore, it is intuitive that clients with similar data distributions should have similar locations of their critical parameters. To demonstrate this intuition, as shown in Figure 1, we conduct an experiment with 15 clients. Figure 1(a) shows the overlap ratio of the locations of critical parameters for any two clients. Figure 1(b) shows the data distribution for 15 clients. Take client 0 as an example. It has the same data distribution as client 1 and class overlap with clients 10 and 14, so their critical parameter location overlap ratio is high. Also, client 0 has a high overlap ratio with clients 8 and 9 due to the similarity of 'Automobile' and 'Truck' data. The result of the experiment is consistent with our intuition. Therefore, in our proposed FedCAC, we utilize the overlap ratio of critical parameter locations to indirectly reflect the client data distribution similarity.

*Corresponding author

B. Visualization of data partitioning in Dirichlet non-IID scenarios

To facilitate intuitive understanding, we utilize 20 clients on the 10-classification dataset to visualize the data distribution of clients with different α values. As shown in Figure 2, the horizontal axis represents the client ID, and the vertical axis represents the data class label index. Red dots represent the data assigned to clients. The larger the dot is, the more data the client has in this class. When α is small (e.g., $\alpha = 0.01$), the overall data distributions of clients vary greatly. However, the variety of client data distribution is low, and it is easy to have clients with very similar data distributions. As the α increases, the difference in data distribution among clients gradually decreases while the variety of client data distribution increases.



(a) $\alpha = 0.01$

(b) $\alpha = 0.1$

(c) $\alpha = 0.5$

(d) $\alpha = 1.0$

Figure 2. Visualization of data partitioning in Dirichlet non-IID scenarios with different α.

C. Additional Comparison with SOTA methods

Parallel with our work, we find two PFL works pFedGate [1] and perFedMask [2] that also propose to use mask matrices. To further enhance the persuasiveness of our work, we add experiments to compare with them.

Since our primary focus is on accuracy, we choose hyperparameter settings that yield the best accuracy for each method. For instance, we set $s = 1$ in pFedGate and $\nu = 1$ in PerFedMask. Table 1 presents the test accuracy achieved by the different methods. Notably, FedCAC outperforms the other methods in terms of accuracy. This improvement can be attributed to FedCAC's ability to facilitate more refined collaboration by considering data distribution similarity and parameter sensitivity, thereby enhancing robustness in non-IID scenarios.

Furthermore, we would like to clarify that although all aforementioned methods utilize masks, their purposes and functionalities differ significantly. While pFedGate utilizes masks for personalized model adaptation and PerFedMask employs masks to freeze layers for computation reduction, our method utilizes masks to control parameter collaboration mode and represent client data distribution.

Methods	CIFAR-10	CIFAR-100
pFedGate (ICML 2023)	89.15 ± 0.76	92.30 ± 0.72
PerFedMask (ICLR 2023)	89.20 ± 0.65	92.12 ± 0.56
FedCAC	$\mathbf{89.77 \pm 1.14}$	$\mathbf{93.05 \pm 0.90}$

Table 1. Comparison results under Pathological non-IID.

References

[1] Daoyuan Chen, Liuyi Yao, Dawei Gao, Bolin Ding, and Yaliang Li. Efficient personalized federated learning via sparse model-adaptation. *arXiv preprint arXiv:2305.02776*, 2023. 2

[2] Mehdi Setayesh, Xiaoxiao Li, and Vincent WS Wong. Perfedmask: Personalized federated learning with optimized masking vectors. In *The Eleventh International Conference on Learning Representations*, 2022. 2