Layer-wised Model Aggregation for Personalized Federated Learning

Supplementary Material

A. Model Architecture And Training Details

In this section, we provide the details of our model and training process on EMNIST [1], FashionMNIST [9], CIFAR10 [4] and CIFAR100 [4].

A.1. Model Architecture

Figure 1 shows the model structure of the hypernetwork used in our paper. First, the hypernetwork uses an embedding layer to generate and update the embedding vector v_i. Then, the embedding vector v_i passes through several public fully connected layers and generates an intermediate feature. Finally, the intermediate feature is sent to n fully connected layers which correspond to n layers of local clients' models. These final fully connected layers output the aggregation weights for each layer of client i respectively.

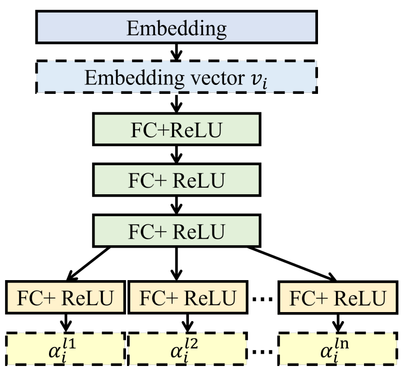


Figure 1. The model structure of the hypernetwork for client i

A.2. Training Details

In all experiments, we use SGD optimizer and cross-entropy loss, the learning rate is 0.01 for CIFAR10 and CIFAR100 while 0.005 for EMNIST and FashionMNIST, the batch size is 32. The local epochs is 20 for 100 clients case and 10 for 10 clients case. In the comparison experiments with state-of-the-art baselines, we set 600 communication rounds for 10 clients case and 2500 communication rounds for 100 clients case, the participation ratio is 100% and 10% respectively. For pFedMe, we set the personal learning rate as 0.01, β as 1 and λ as 15. For FedFomo, the number of local models that the server sends to one client in each communication round is 5.

B. Additional Results

In this section, we provide more extensive experimental results to compare the training performance of pFedLA and HeurpFedLA with Local Training, FedAvg [6], Per-FedAvg [3], pFedMe [8], pFedHN [7], FedBN [5], FedRep [2], FedFomo [10]. Besides, we present additional work on key hyperparameters of pFedLA to give further insight into our method's functionality and efficiency to parameters. We consider the effect of partial aggregation on training performance of pFedLA.

B.1. The Training Performance on Different non-IID Settings

We compare pFedLA with state-of-the-art baselines on four datasets: EMNIST, FashionMNIST, CIFAR10 and CIFAR100. There are two non-IID data settings: 1) each client is randomly assigned four classes (twelve classes per client in CIFAR100) with the same amount of data on each class; 2) each client contains all classes, while the data on each class is not uniformly distributed. Figure 2, 3, 4, 5 shows the empirical convergence results of pFedLA along with other baselines. Specifically, we focus on the changes of average test accuracy of these algorithms in each communication round. It is obvious that pFedLA converges to higher average test accuracy on all four datasets than baselines. This phenomenon validate the effectiveness of the parameterized layer-wised model aggregation on pFL training.

B.2. Effect of Partial Aggregation

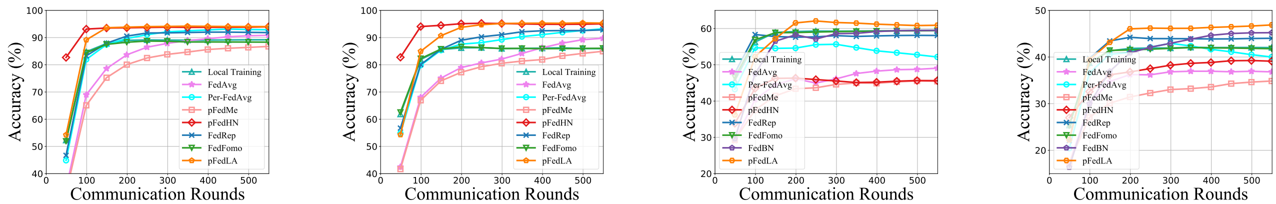
To address the unreliable communication and computation environment challenges in personalized federated learning, we conduct experiments to explore the effect of partial aggregation on training performance of pFedLA. Specifically, in large-scale FL system, we only aggregate M ($M \ll N$) clients with the largest weight in each communication round. Table 1 shows the final model accuracy for different values of M. The total number of clients is 100 and the participation ratio is 10%. It can be noticed that a proper M can achieve better performance, which demonstrates that the value of M should be carefully designed.

Dataset \ # Value of M	4	5	7	9
CIFAR10 (%)	72.23	73.15	**74.27**	71.46
EMNIST (%)	95.25	**96.34**	93.99	92.90

Table 1. The Model Accuracy under Different setting of M.

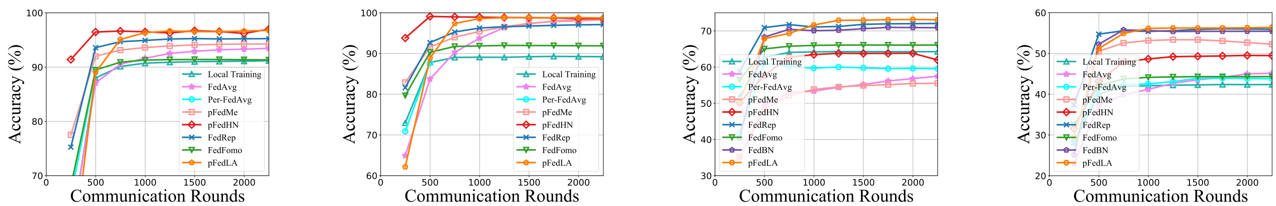
B.3. Visualization of the Aggregation Weight

In section 4.3, we have discussed the relationship between the aggregation weights and the data similarities



(a) EMNIST (b) FashionMNIST (c) CIFAR10 (d) CIFAR100

Figure 2. Performance of pFedLA compared with baselines for 10 clients case on different datasets, non-IID_1.



(a) EMNIST (b) FashionMNIST (c) CIFAR10 (d) CIFAR100

Figure 3. Performance of pFedLA compared with baselines for 100 clients case on different datasets, non-IID_1.

among clients. Figure 4 (in main paper) shows the visualization of the aggregation weights in FC1 layer on EMNIST and FashionMNIST, FC3 layer on CIFAR10 and CIFAR100. We additionally report heatmap on other layers in Figure 6. The experimental setup is consistent with that in the main paper. It can be noticed that the self-weights of each client still have the highest values. The weights among close clients with consecutive IDs are still larger than those of the distant clients. These results further verify that pFedLA can exploit the inter-similarities among heterogeneous clients.

C. Analysis of Additional FLOPs of pFedLA

In pFedLA, each client maintains a hypernetwork to produce the personalized aggregation weights. The training of the hypernetworks may bring additional computation overhead. Table 2 shows the FLOPs of the model on a client and a hypernetwork on the server. When compared with a hypernetwork, the forward propagation on the client is even larger. What's more, it is just the computation needed for one pass on one sample. In each communication round, the client models have to train several local steps on the whole private data samples, while a hypernetwork only needs to calculate the aggregation weights once in each round. Therefore, the additional computation of pFedLA is acceptable.

References

[1] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. Emnist: Extending mnist to handwritten letters. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2921–2926. IEEE, 2017. 1

[2] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *Proceedings of the 38th International Conference on Machine Learning, ICML*, volume 139, pages 2089–2099. 1

[3] Alireza Fallah, Aryan Mokhtari, and Asuman E. Ozdaglar. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. In *Advances in Neural Information Processing Systems, NeurIPS*, 2020. 1

[4] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 1

[5] Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. In *Proceedings of the 9th International Conference on Learning Representations, ICLR*. OpenReview.net, 2021. 1

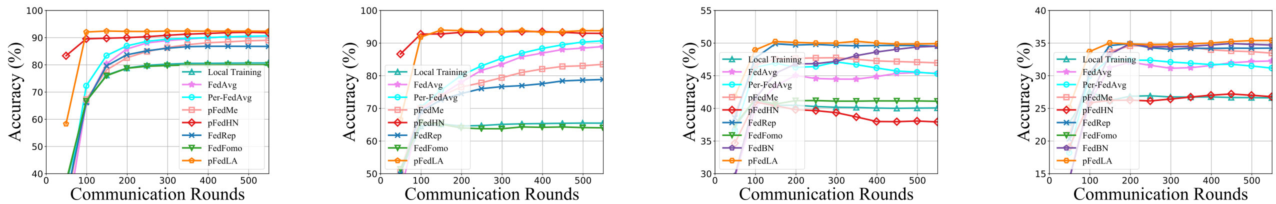
[6] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*, pages 1273–1282. PMLR, 2017. 1

[7] Aviv Shamsian, Aviv Navon, Ethan Fetaya, and Gal Chechik. 1

[8] Canh T Dinh, Nguyen Tran, and Tuan Dung Nguyen. Personalized federated learning with moreau envelopes. *Proceedings of Advances in Neural Information Processing Systems, NeurIPS*, 33, 2020. 1

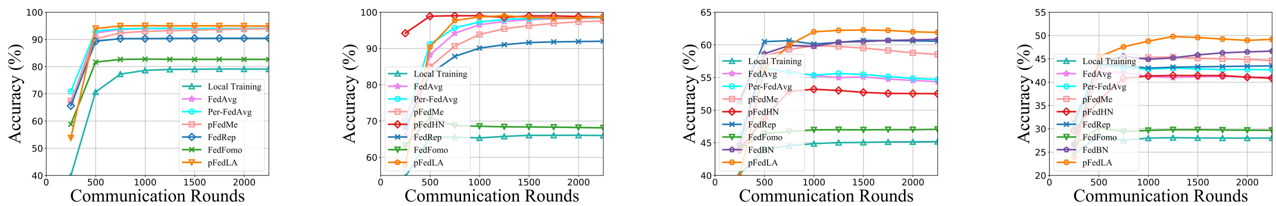
[9] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. 1

[10] Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, and Jose M. Alvarez. Personalized federated learning with first order model optimization. In *Proceedings of the 9th International Conference on Learning Representations, ICLR 2021*, 2021. 1



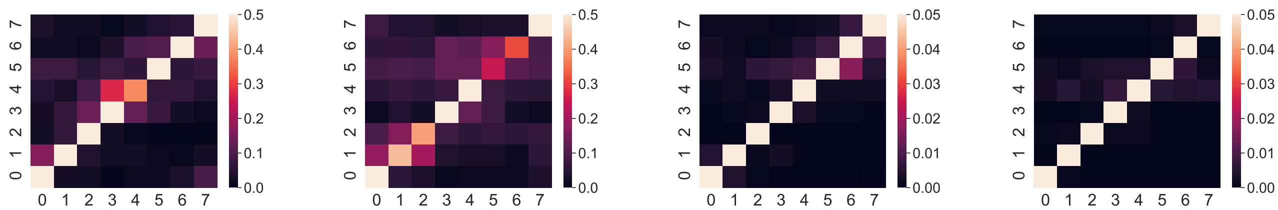
(a) EMNIST (b) FashionMNIST (c) CIFAR10 (d) CIFAR100

Figure 4. Performance of pFedLA compared with baselines for 10 clients case on different datasets, non-IID_2.

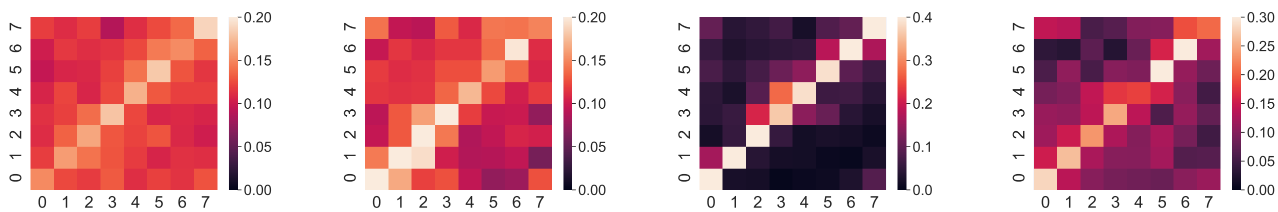


(a) EMNIST (b) FashionMNIST (c) CIFAR10 (d) CIFAR100

Figure 5. Performance of pFedLA compared with baselines for 100 clients case on different datasets, non-IID_2.



(a) EMNIST (FC3) (b) FashionMNIST (FC3) (c) CIFAR10 (FC1) (d) CIFAR100 (FC1)



(e) EMNIST (Conv1) (f) FashionMNIST (Conv1) (g) CIFAR10 (Conv2) (h) CIFAR100 (Conv2)

Figure 6. The visualization of the aggregation weights in a specific layer on EMNIST, FashionMNIST, CIFAR10 and CIFAR100. X-axisand y-axis show the IDs of clients.

Model \ Dataset	EMNIST	FashionMNIST	CIFAR10	CIFAR100
Client Model	2266.45K	2266.45K	4749.90K	4765.02K
Hypernetwork	101.23K	101.23K	117.29K	117.29K

Table 2. FLOPs of the client model and hypernetwork.