

# Supplementary Materials: Decoupling General and Personalized Knowledge in Federated Learning via Additive and Low-rank Decomposition

ANONYMOUS AUTHORS

## 1 PARTIAL CLIENT PARTICIPATION

In our previous experiments, we assume all clients participate in FL training in each round. However, some clients may be offline due to reasons such as unstable communication links. This is the partial client participation problem that is common in FL. In this section, we evaluate the robustness of FedDecomp to this problem. We consider the scenarios where 90%, 70%, and 50% clients participate in each round and carry experiments on CIFAR-10, CIFAR-100, and Tiny Imagenet with  $\alpha = 0.1$ .

The results are illustrated in Table 1. As we can see, in all scenarios, partial client participation does not significantly affect accuracy compared to all client participation. This is attributed to FedDecomp’s effectively separating general knowledge and client-specific knowledge, and the effect of non-IID is reduced through alternate training. Ensure that client collaboration is not significantly affected by outline clients in each round.

## 2 ADDITIONAL EXPERIMENTS ON LARGER DATASETS

While the current mainstream FL work focuses on the algorithm’s performance on small image datasets, in this section, we further verify the performance of FedDecomp on larger datasets as well as other modality datasets.

Specifically, we conduct additional experiments on both a larger image dataset and a natural language processing (NLP) dataset. For the larger image dataset, we select a subset from ImageNet, consisting of 400 classes with a total of 80,000 samples. We utilize the ResNet-10 model architecture, with each client having 2,000 training samples generated following the Dirichlet distribution with  $\alpha = 0.1$ . For the NLP dataset, we opt for AG\_NEWS, a text 5-classification dataset with 120,000 samples. We employ the Transformer model architecture, with each client having 3,000 training samples generated following the Dirichlet distribution with  $\alpha = 1.0$ . Additionally, for the Transformer model, we apply model decomposition to the weights in the self-attention modules and fully connected weights in the classifier module.

Table 2 displays the test accuracy results for these two datasets. It’s evident that FedDecomp consistently outperforms other state-of-the-art methods on both datasets.

## 3 WHETHER FEDDECOMP SACRIFICES SOME CLIENTS’ ACCURACY

In previous experiments, we demonstrate the improvement of the averaged accuracy of all clients. In this section, we focus on the individual improvement for each client and verify whether FedDecomp sacrifices some clients’ accuracy. We plot each client’s accuracy in FedDecomp, FedAvg, and Local (i.e., each client trains the model locally without collaboration) methods in the Dirichlet non-IID scenario with  $\alpha = 0.1$ . The results are shown in Fig. 1.

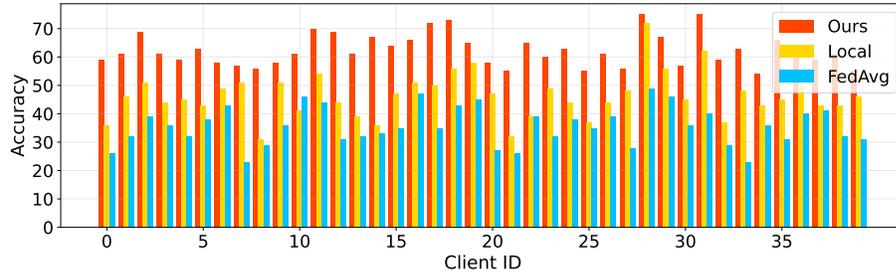
Notice that the accuracy of all clients in the FedDecomp method is higher than that in the FedAvg and Local methods, affirming that the use of FedDecomp does not lead to any deterioration in individual client performance.

Table 1. The effect of partial client participation.

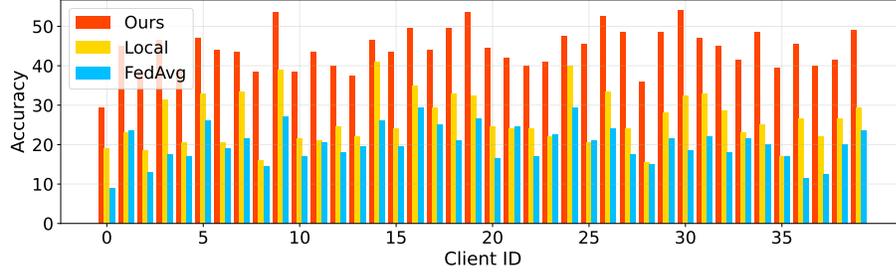
Datasets	100%	90%	70%	50%
CIFAR-10	85.47±2.06	85.38±1.62 (-0.09)	85.25±1.67 (-0.22)	85.36±1.67 (-0.11)
CIFAR-100	63.65±0.53	63.01±0.10 (-0.64)	63.21±0.18 (-0.44)	63.13±1.05 (-0.52)
Tiny	44.22±0.55	44.13±0.70 (-0.09)	44.10±0.26 (-0.12)	43.99±0.62 (-0.23)

Table 2. Comparison results on larger datasets.

Datesets	FedAvg	FedPer	FedRoD	FedDecomp
AG_NEWS	89.36	90.76	91.38	<b>91.79</b>
ImageNet-Subset	18.55	29.37	32.45	<b>35.67</b>



(a) CIFAR-100



(b) Tiny Imagenet

Fig. 1. Test accuracy of each client in Dirichlet non-IID scenario with  $\alpha = 0.1$ .

#### 4 VISUALIZATION OF DATA PARTITIONING IN DIRICHLET NON-IID

To facilitate intuitive understanding, we utilize 20 clients on the 10-classification and 50-classification datasets to visualize the data distribution of clients with different  $\alpha$  values. As shown in Figure 2, the horizontal axis represents the data class label index, and the vertical axis represents the client ID. Red dots represent the data assigned to clients. The larger the dot is, the more data the client has in this class. When  $\alpha$  is small (e.g.,  $\alpha = 0.1$ ), 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

Author's address: Anonymous Authors.

Manuscript submitted to ACM

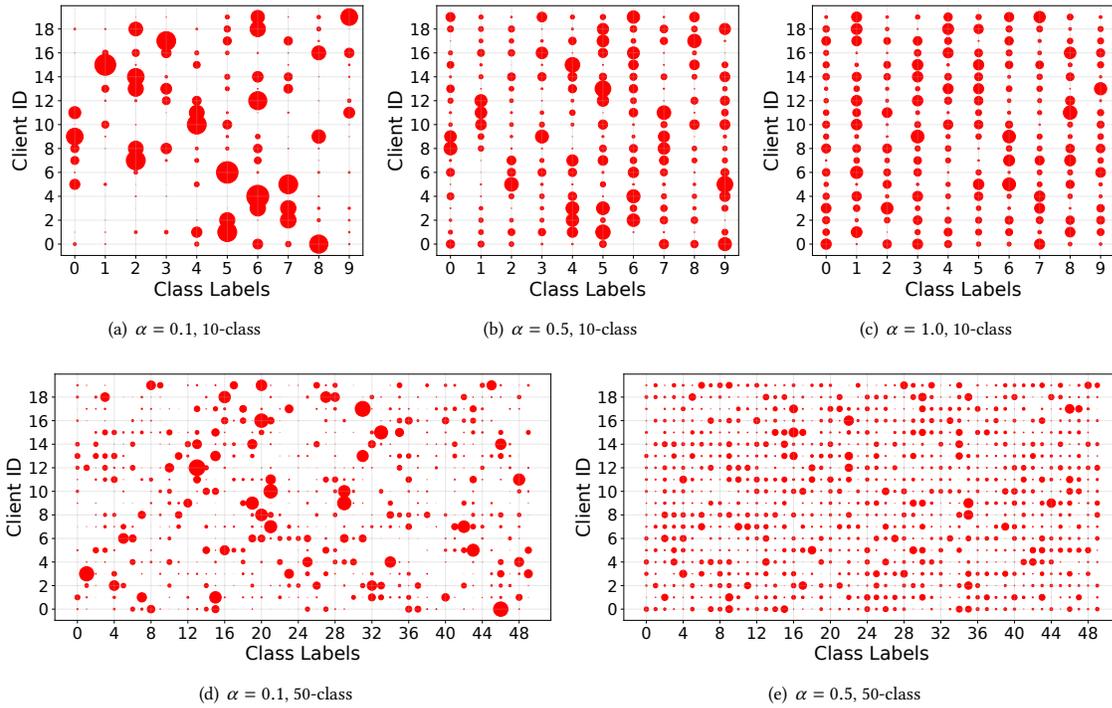


Fig. 2. Visualization of data partitioning in Dirichlet non-IID scenarios with different  $\alpha$ .

data distributions. As the  $\alpha$  increases, the extent of class imbalance within each client’s dataset gradually diminishes, consequently leading to more difficult local tasks (i.e., the number of classes involved and a reduction in the number of samples available for each class). Concurrently, the dissimilarity in data distribution among different clients gradually diminishes, while the diversity in client data distribution widens. Furthermore, comparing the 10-classification dataset and the 50-classification dataset, it can be seen that under the same  $\alpha$  value, when the number of dataset classes increases, the difference of client data distribution becomes larger, and the diversity of client data distribution increases. It becomes more difficult to extract general knowledge among clients.

In summary, the Dirichlet non-IID configuration proves to be a potent approach for assessing the performance of PFL methods across a spectrum of intricate and diverse non-IID scenarios.