Exploiting Features and Logits in Heterogeneous Federated Learning

Yun-Hin Chan and Edith C.-H. Ngai

Department of Electrical and Electronic Engineering, The University of Hong Kong

香 港 大 學
THE UNIVERSITY OF HONG KONG

Background

Federated learning (FL) facilitates the management of edge devices to collaboratively train a shared model while maintaining training data local and private. However, a general assumption in FL is that all edge devices are trained on the same machine learning model, which may be impractical considering diverse device capabilities. For instance, less capable devices may slow down the updating process because they struggle to handle large models appropriate for ordinary devices. A system containing clients with heterogeneous capabilities is referred to as system heterogeneity, which is one of the most critical challenges in FL.

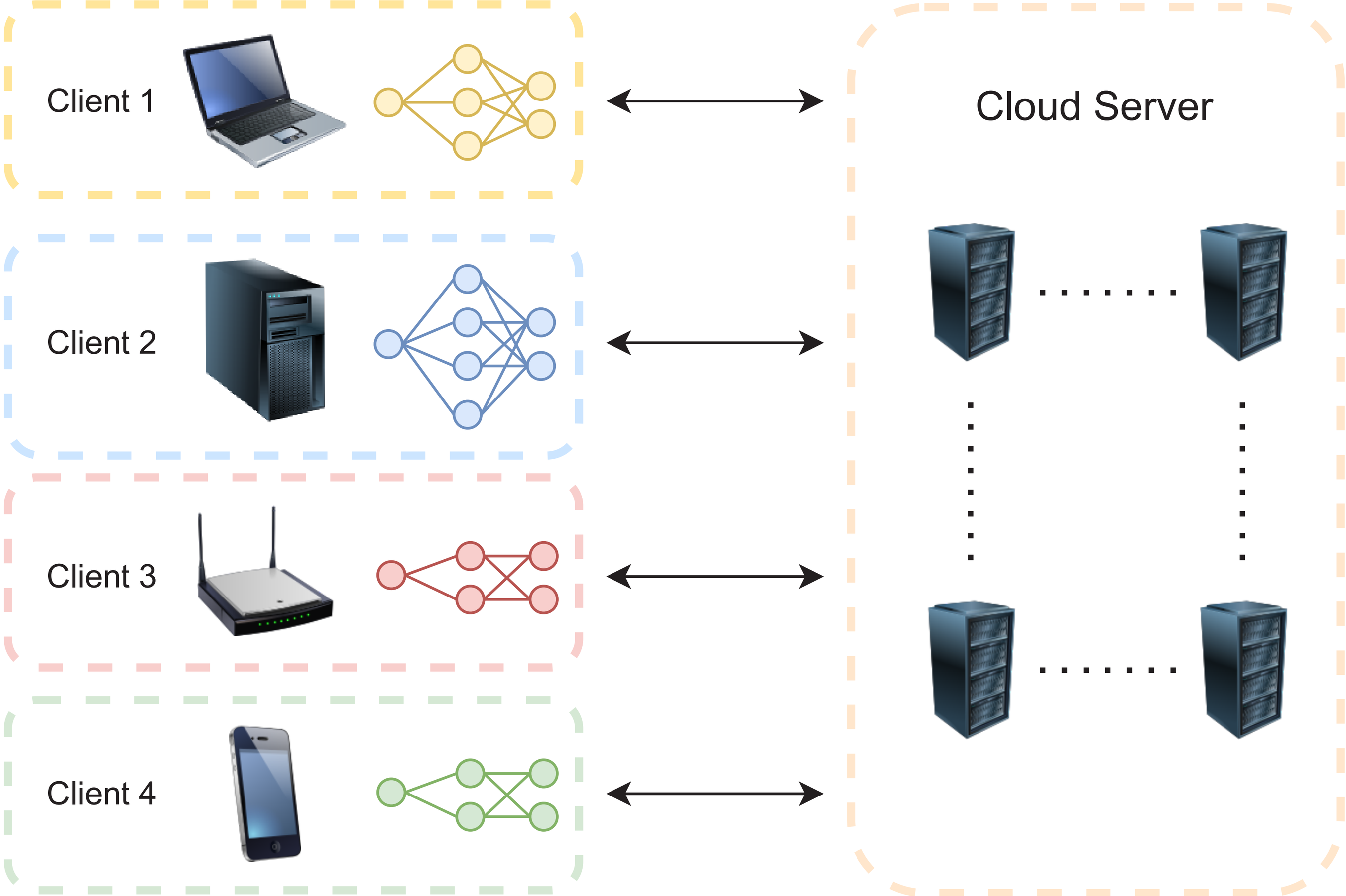

Figure 1: The illustration of system heterogeneity in FL in IoT.

Methods

We proposed two novel data-free knowledge distillation methods, called Felo and Velo, to support system heterogeneity in FL.

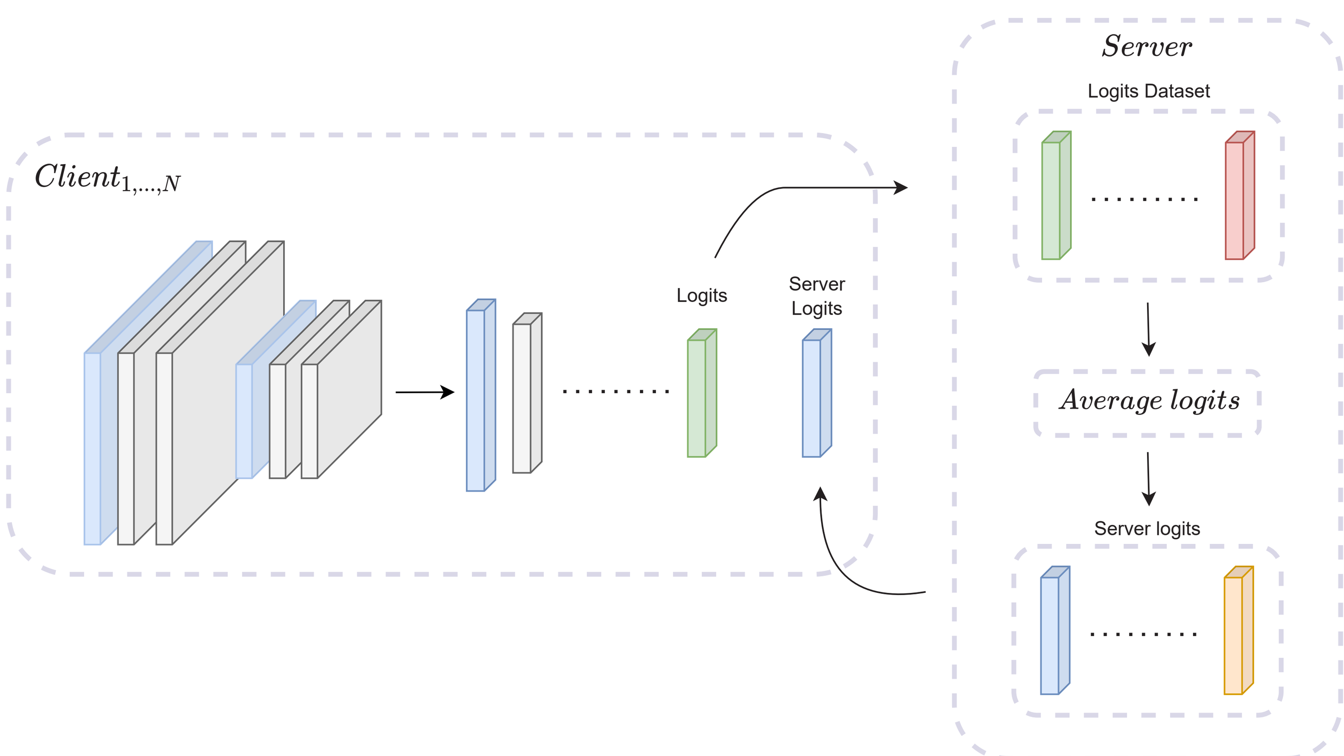

Figure 2: Architecture design of Felo.

In Figure 2, the clients train their models based on their private data, and collect mid-level features and logits, which are then transmitted to a server. The server aggregates this information according to their class labels. Finally, this server sends these aggregated features and logits back to clients, which will be utilized to train the client models.

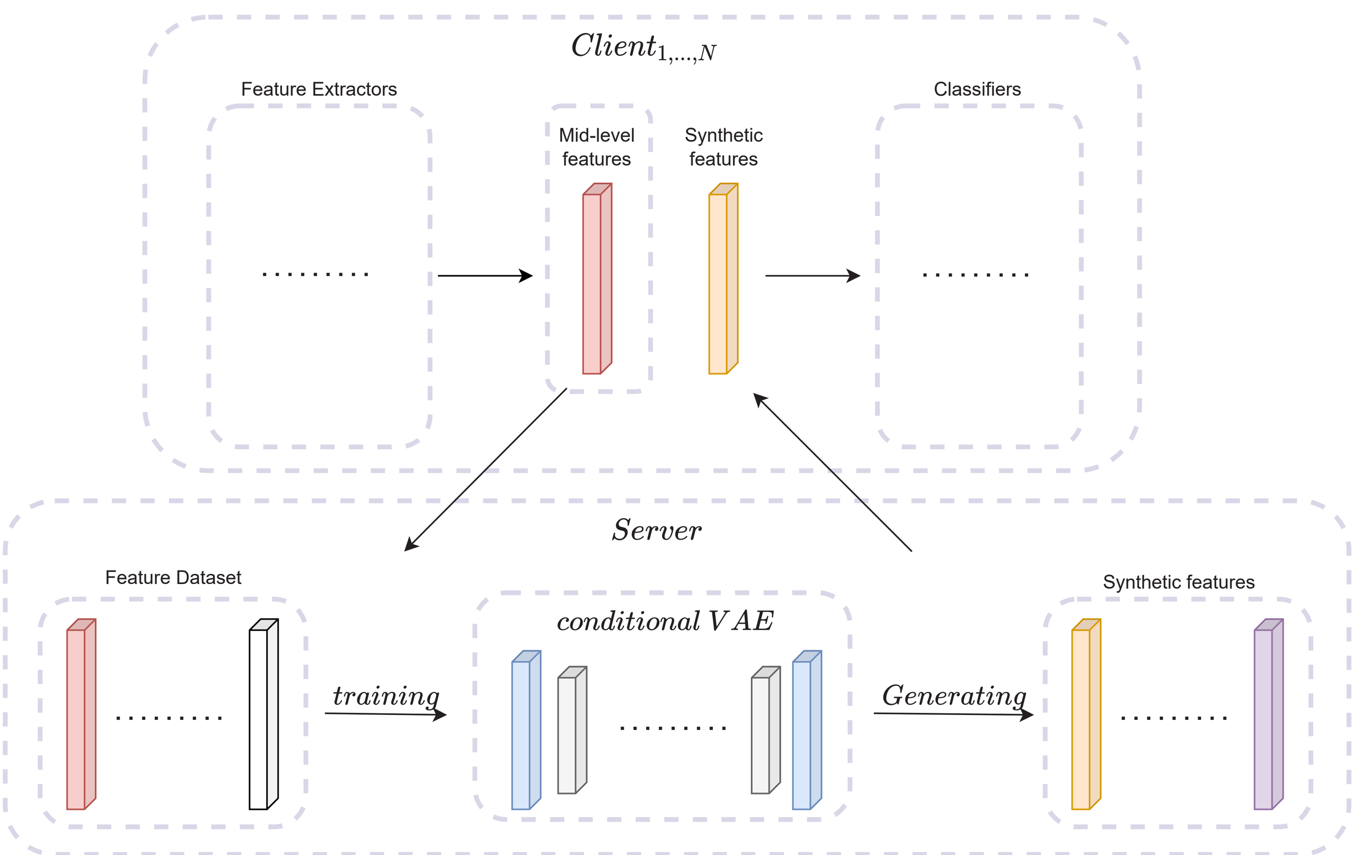

Figure 3: Architecture design of Velo.

In Figure 3, the server uses mid-level features from the feature dataset to train the CVAE in Velo. The rest of the process is the same as Felo.

Experimental Results

Table 1: Accuracy of FedAvg, Felo and Velo in CIFAR-10

Method	CIFAR-10	
	iid	non-iid
FedAvg	84.582±0.26%	59.229±0.22%
Felo	84.451±0.41%	60.357±0.52%
Velo	**85.077±0.32%**	**60.882±0.40%**

We have conducted experiments to compare the accuracy of Felo and Velo with FedAvg (McMahan et al., 2017). The results are shown in Table 1 on iid and non-iid data from the CIFAR-10 dataset. Velo obtains the highest accuracy among the three methods in both iid and non-iid settings.

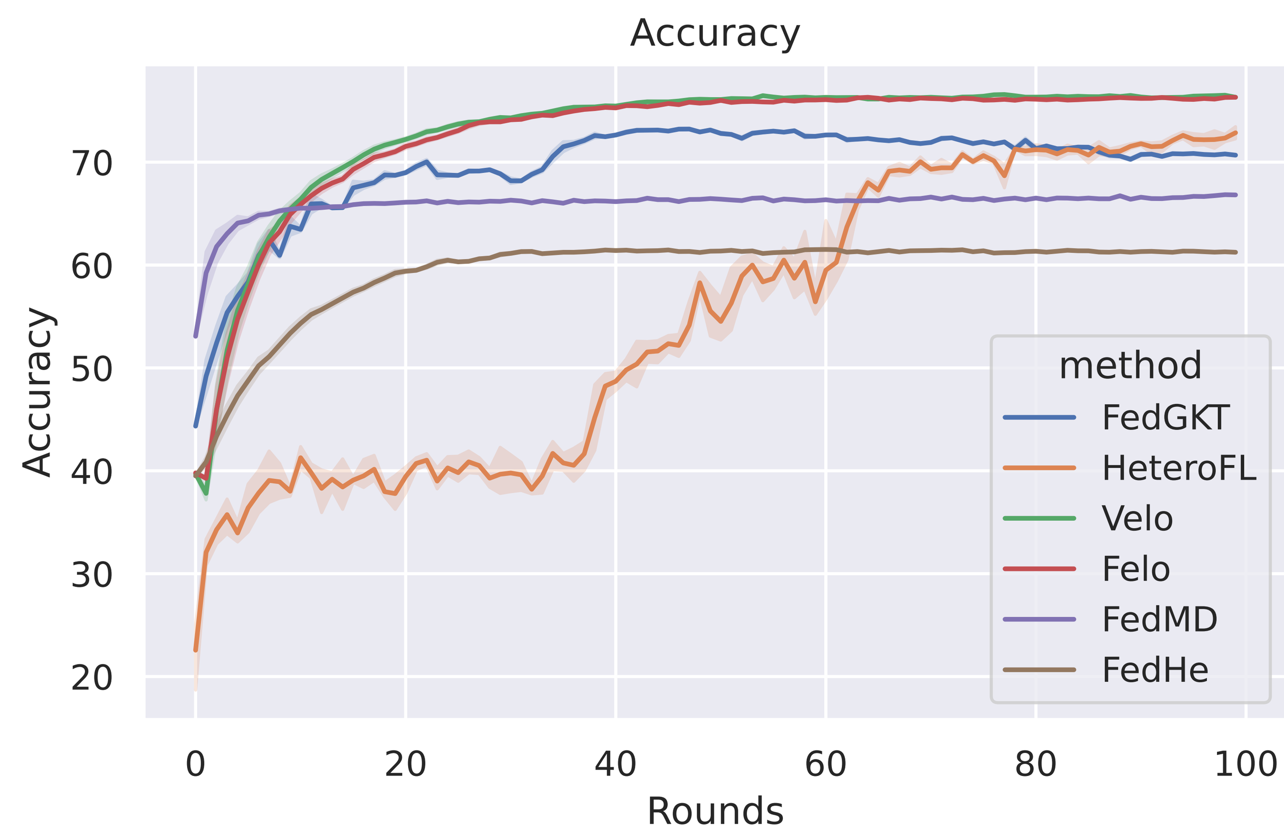

Figure 4: Accuracy of methods in iid CIFAR-10.

From Figure 4, Velo achieves the highest accuracy of 76.61% in iid CIFAR-10. The second-best accuracy is achieved by Felo, which is 76.35%. HeteroFL obtains the third-best performance with the accuracy of 73.56%. Moreover, FedGKT achieves 73.27% and FedMD attains 66.88% in model accuracy. FedHe(Chan & Ngai, 2021) achieves 61.56% in this experiment.

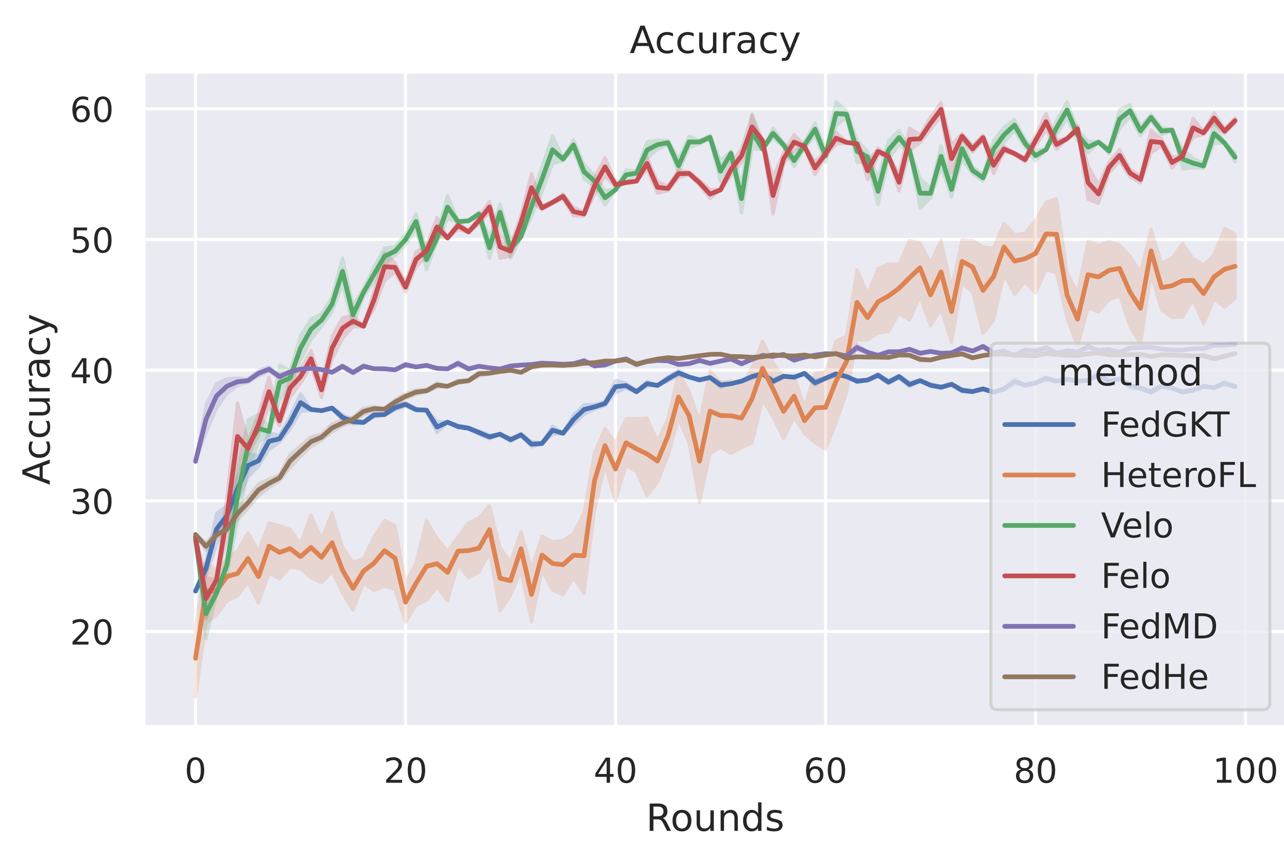

Figure 5: Accuracy of methods in non-iid CIFAR-10.

In Figure 5, in non-iid CIFAR-10, the best model accuracy still comes from Velo, 60.56%, and the second one is 60.48% from Felo. The following algorithms are HeteroFL and FedMD, obtaining model accuracy of 55.23% and 42.01%, respectively. The accuracy of FedHe is 41.35%, while the worst is FedGKT with only 40.01%.

References

Chan, Y.-H., & Ngai, E. (2021). FedHe: Heterogeneous models and communication-efficient federated learning. *IEEE International Conference on Mobility, Sensing and Networking (MSN 2021)*.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273–1282).