A Naive Baseline in Heterogeneous Federated Learning

Durmus Alp Emre Acar
Boston University

Standard federated learning settings assume model architectures on the devices to be the same which may be unrealistic. [1] proposes HeteroFL, for federated training of heterogeneous architectures.

Naive Baseline. We compare HeteroFL against a naive baseline. We group devices that share the same architecture and perform federated learning only over that group. In this note, we tabulate performance of such naive baselines and compare it against [1] in the setting of two different architectures (simple and complex). Our naive baseline trains a complex model based on running FedAvg algorithm [2] and protocol exclusively on the complex devices. In an analogous fashion we train the simple model.

We report test results on CIFAR10 dataset based on an IID split. There are 50 complex and 50 simple models and the activation rate is 10%. To construct the simple model, we shrink the CNN layers using 0.0625 complexity level. We use the available HeteroFL code for ResNet experiment and our results appear to validate reported test accuracy of [1]. Results for the two different experiments are reported in 1 and 2.

Table 1: Complex model: Neural-Net consisting of three convolutional layers and includes group normalization.

Algorithm	Server Complex Accuracy Architecture 'a'	Server Simple Accuracy Architecture 'e'
Decoupled FedAvg on 50 Complex Clients	79.9	NA
Decoupled FedAvg on 50 Simple Clients	NA	66.6
HeteroFL	79.6	64.8

Table 2: Complex model: ResNet model with an sBN module as in [1].

Algorithm	Server Complex Accuracy Architecture 'a'	Server Simple Accuracy Architecture 'e'
Decoupled FedAvg on 50 Complex Clients	89.1	NA
Decoupled FedAvg on 50 Simple Clients	NA	77.4
HeteroFL	89.3	74.3

Discussion. The performance of the naive baseline is on-par with HeteroFL across different architectures for the complex model. On the other hand, decoupled training for simple models outperforms HeteroFL. Our results raise the question that leveraging federated learning for training heterogeneous device models may result in loss of accuracy. In particular, the benefits from accessing more devices (and so more data), is outweighed by the difficulty in exploiting/transferring knowledge across different architectures.

References

[1] Enmao Diao, Jie Ding, and Vahid Tarokh. Hetero{fl}: Computation and communication efficient federated learning for heterogeneous clients. In *International Conference on Learning Representations*, 2021.

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