Quantifying the Impact of Label Noise on Federated Learning
Shuqi (Griffith) Ke 1  Chao Huang 2  Xin Liu 3
1shuqike@link.cuhk.edu.cn  2fchhuang@ucdavis.edu  3xinliu@ucdavis.edu

Problem

- **Label noise** has been identified in many widely used datasets. The cause of label noise could be human error.
- **Federated Learning** (FL) is a distributed machine learning paradigm where clients collaboratively train a model using their local (human-generated) datasets.

How does label noise affect FL generalization?

Observation

Label noise **linearly** degrades FL performance by reducing the test accuracy of the global model.

![Linear regression on the global model accuracy under label noise. Two clients trained the global model with FedAvg.](image1)

**Method**

Consider an FL process with a $L$-layer neural network $f : \mathbb{R}^{d_i} \times W \rightarrow \mathbb{R}^{d_o}$ as its global model. Then its path-norm increases at most polynomially,

$$\|f(\cdot; \theta(t))\|_{\text{pnp}} = O(t^{L+1}E^{L+1/2})$$

where $t \leq R$ denotes the number of communication rounds and $E$ denotes the local training time.

If we consider a generic decentralized algorithm, we have

$$\|f(\cdot; \theta(t))\|_{\text{pnp}} = O(e^{C't(L+1)}E^{L+1/2})$$

where $C'$ is a constant independent of $t, L, E$.

**Theoretic analysis**

**Formulation**

![Diagram of the model architecture](image2)

Figure 3. We consider a classification task and ReLU networks.

**Method**

Consider any distributed/decentralized learning algorithm with a neural network with an arbitrary structure for a classification task of $C$ classes under label noise and cross-entropy loss, then

$$G(W) \leq \Omega \cdot \mathbb{E}_X \left\{ \sum_{i=1}^C \sum_{j=1}^{\tilde{N}_i} |\Pr_{\tilde{X}}(Y = i|X) - \Pr_{X}(Y = i|X)| \right\}$$

If we use a ReLU network as our model in the FL task, then $\Omega = \|f(\cdot; \theta(t))\|_{\text{pnp}}$.

**Contributions**

- The first quantitative study that analyzes the impact of label noise on FL.
- An analysis on path-norm growth under distributed/decentralized setting.
- A linear upper bound on the generalization error of the global model under label noise, which is consistent with the empirical results.

**Discussion**

- Improving theoretical bounds: We prove a linear upper bound for the generalization error, and the bound is consistent with numerical results. However, the upper bound can be loose. One can provide a lower bound or improve the upper bound by making more restrictive assumptions.
- More comprehensive experiments: Our experiments use a small number of clients, which applies to cross-silo FL. In future research, we plan to study the impact of label noise with a larger number of clients (e.g., as in cross-device FL).

The AAAI 2023 Workshop on Representation Learning for Responsible Human-Centric AI, Washington DC