Overcoming Noisy and Irrelevant Data in Federated Learning

Tiffany Tuor[1], Shiqiang Wang[2], Bong Jun Ko[3], Changchang Liu[2], Kin K. Leung[1]
Imperial College London [1], IBM T. J. Watson Research Center[2],Stanford Institute for Human-Centered Artificial Intelligence[3]

Motivation:

- Federated learning is an effective way of training a machine learning model from data collected at the tactical edge
- In a coalition, local data collected at the edge are likely to be of diverse types and possibly with noisy labels.
- How to select the subset of data that are relevant for a given federated learning task?

Proposed approach:

Idea: Use a benchmark model trained on a small benchmark dataset, that is task specific, to evaluate the relevance of individual data samples at each clients, and select the data with sufficiently high relevance

Step 1: The benchmark dataset is divided into training $\mathcal{B}_{train}$ and testing $\mathcal{B}_{test}$

Step 2: The benchmark model $\theta_{\mathcal{B}}$ is obtained by training on $\mathcal{B}_{train}$

Step 3: Each client evaluates its own dataset against the benchmark model $\theta_{\mathcal{B}}$ and creates a list of loss values:

$$P_n = \{l(f(x_i, \theta_{\mathcal{B}}), y_i) : \forall (x_i, y_i) \in \mathcal{D}_n\}$$

and the set V which provides a reference distribution of loss values can also be obtained :

$$V = \{l(f(x_i, \theta_{\mathcal{B}}), y_i) : \forall (x_i, y_i) \in \mathcal{B}_{test}\}$$

Step 4: The server merges the lists of the loss values from all clients $P = \cup_{n=1}^{N} P_n$

Step 5: V is used as a mask to find an upper limit of acceptable loss values via a statistical test that compares the distribution of V and P , the threshold of loss value is obtained:

$$\lambda^* = argmin_\lambda \sup_x |F_V(x) - F_P^\lambda(x)|$$

Step 6: Then, each client makes the selection of relevant data locally

$$\mathcal{F}_n = \{(x_i, y_i) \in \mathcal{D}_n : l(f(x_i, \theta_{\mathcal{B}}), y_i) \leq \lambda\}$$

Each client performs stochastic gradient descent on the selected data (batch size adapted to size of $\mathcal{F}_n$)

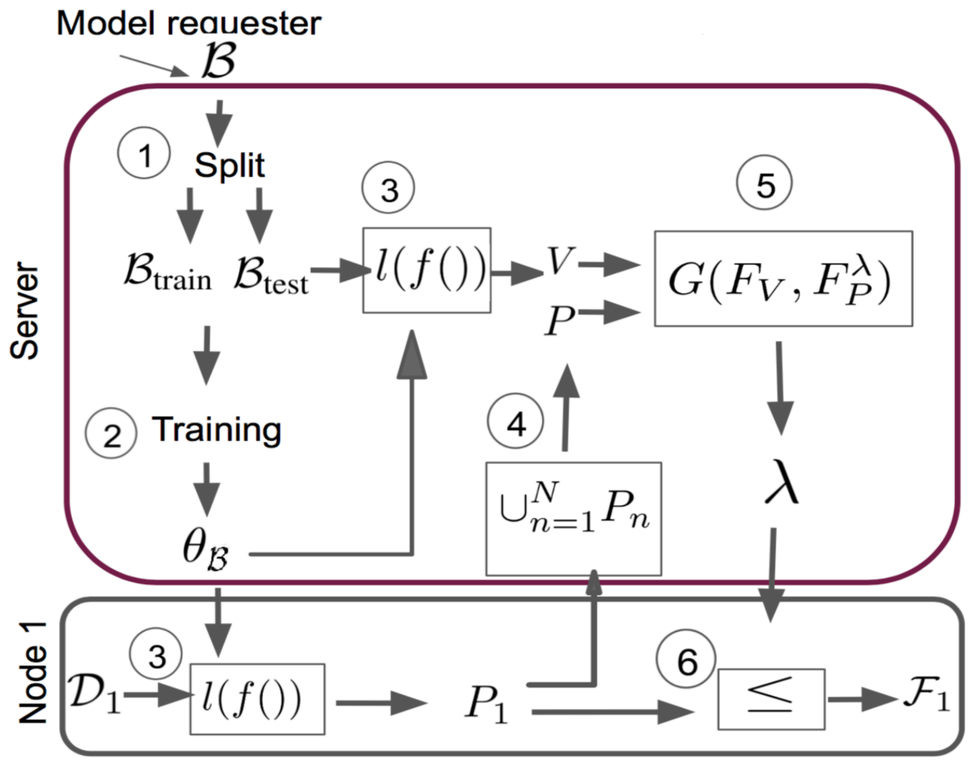
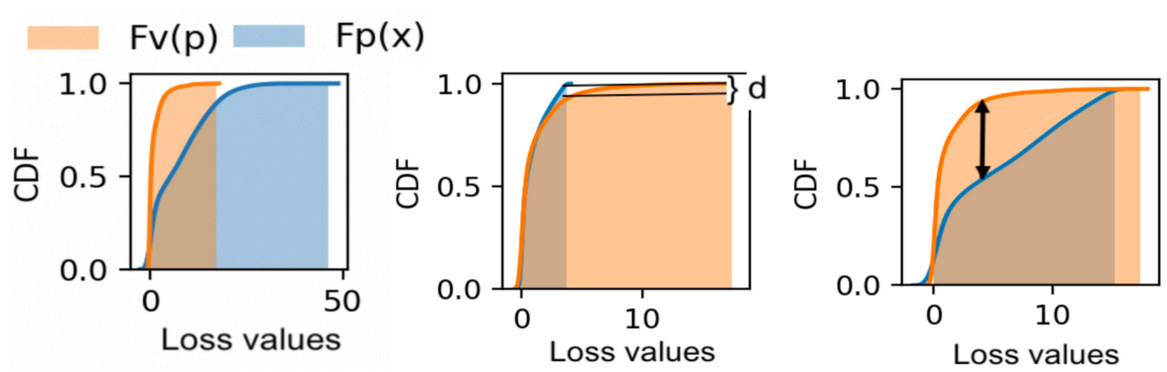


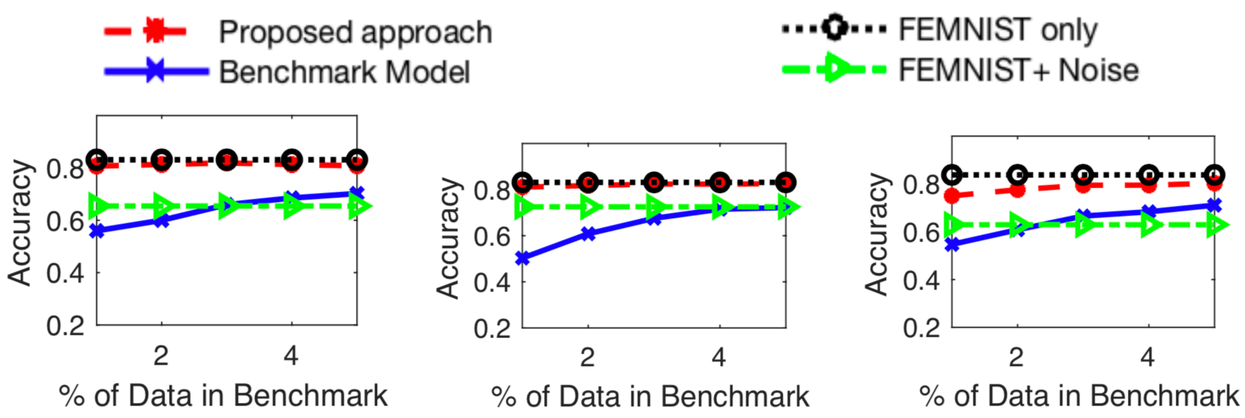
Fig. 1: Data selection procedure.



(a) $F_V(x)$ and $F_P(x)$ (b) $\lambda = 4.8$ (c) $\lambda = 15$

Fig. 2: KS distance computation and optimal λ for $F_V(x)$ and $F_P^\lambda(x)$).

Experiments:



(a) Noise: CHAR-74, Target: FEMNIST (b) Noise: CIFAR-100, Target: FEMNIST (c) Noise: Closed-set, Target: FEMNIST

- Experiments show accuracy achieved for varying amount of benchmark data from 1% to 5% when classifying FEMNIST under different types of noise
- Our approach always performs close to the best case line and also better than the benchmark model and the one trained with the entire noisy dataset → robustness of our approach to both open-set and closed-set noises.
- The performance of the benchmark model increases with the benchmark dataset size while the performance of our approach remains nearly constant.