Overcoming Noisy and Irrelevant Data in Federated Learning

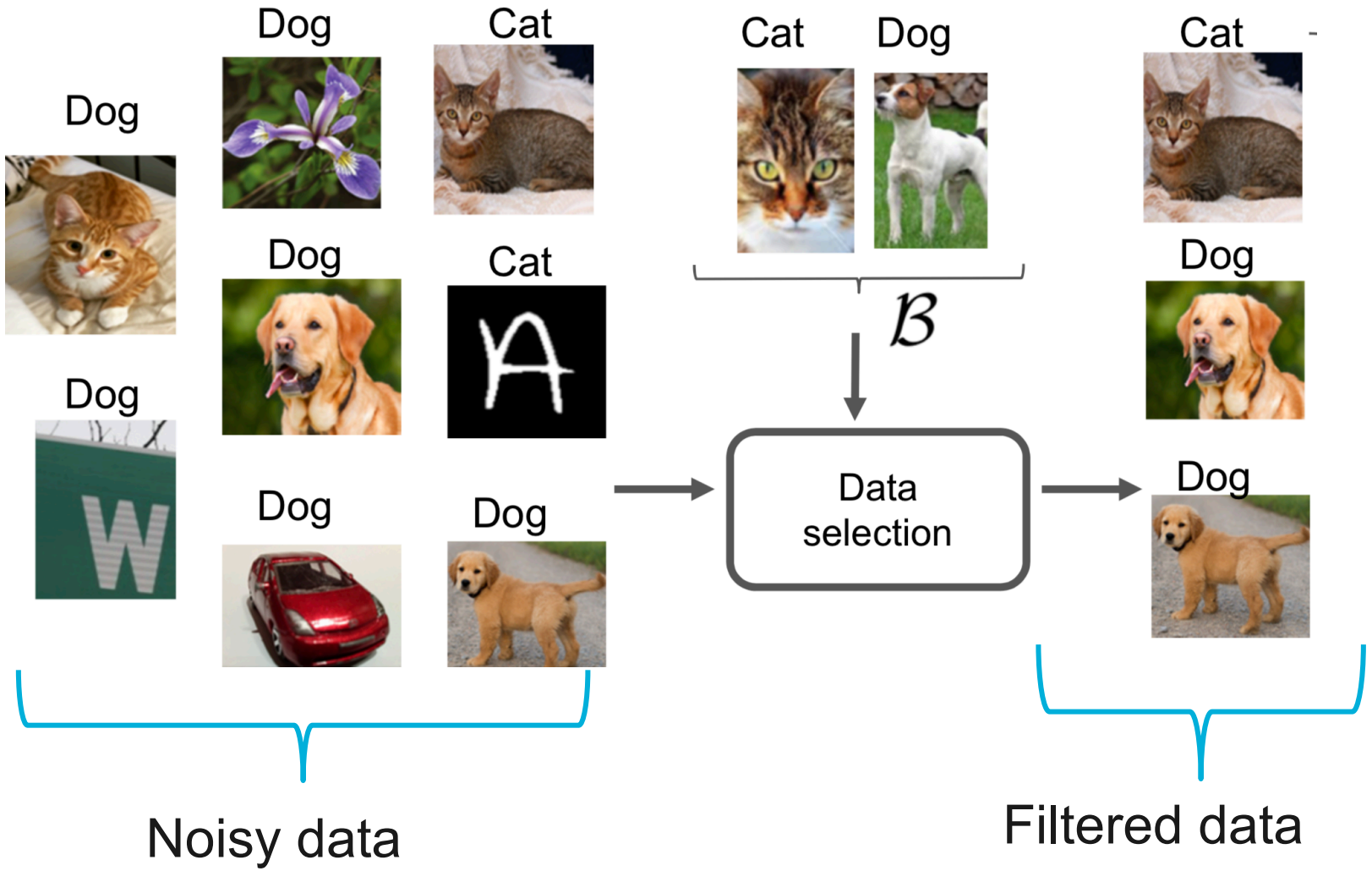
Tiffany Tuor[1], Shiqiang Wang[2], Bong Jun Ko[3], Changchang Liu[2], Kin K. Leung[1]

Imperial College London [1], IBM T. J. Watson Research Center[2],Stanford Institute for Human-Centered Artificial Intelligence[3]

Motivation

- Federated learning allows training of machine learning model from data collected at different clients/locations, without centralizing the raw data.
- Existing approaches focused on training a single global model using pre-defined datasets at client devices.
- In practice, each client can have a large variety of data, possibly with noisy labels, which may or may not be relevant to the given machine learning task and which might impact the accuracy of the global model.
- How to select - in a distributed way - the subset of data that are relevant for a given federated learning task?

Data selection : Overview



Data selection: proposed approach

1) $\mathcal{B}$ is divided into training $\mathcal{B}_{train}$ and testing $\mathcal{B}_{test}$

2) Benchmark model $\theta_{\mathcal{B}}$ is obtained by training on $\mathcal{B}_{train}$

3) Each client evaluates its dataset against $\theta_{\mathcal{B}}$ and creates : $P_n = \{l(f(x_i, \theta_{\mathcal{B}}), y_i) : \forall (x_i, y_i) \in \mathcal{D}_n\}$.

A reference distribution of loss values is also obtained : $V = \{l(f(x_i, \theta_{\mathcal{B}}), y_i) : \forall (x_i, y_i) \in \mathcal{B}_{test}\}$

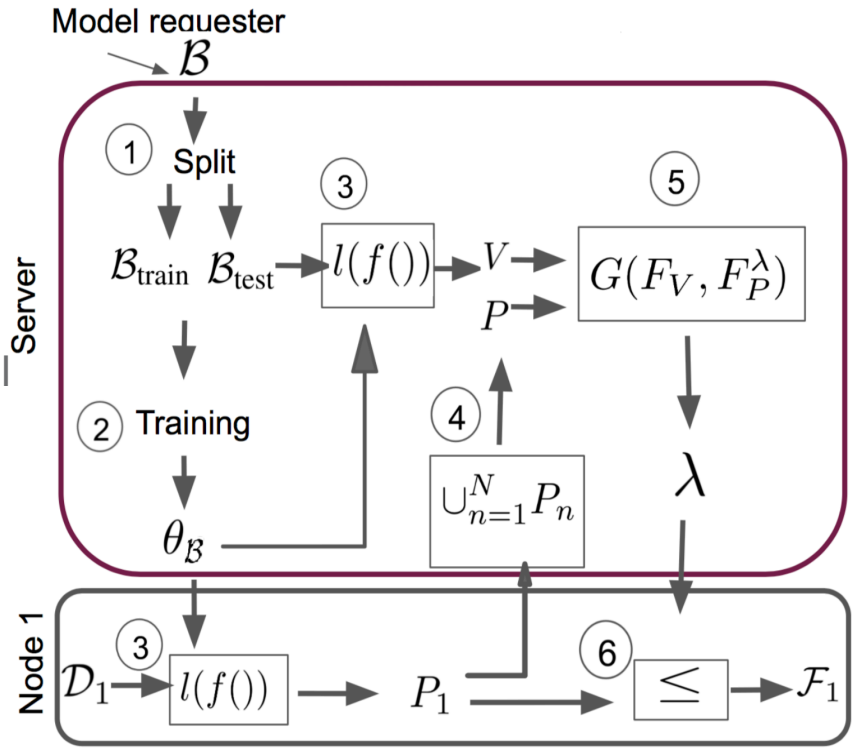
4) The server merges lists of the loss values from all clients

$$P = \cup_{n=1}^{N} P_n$$

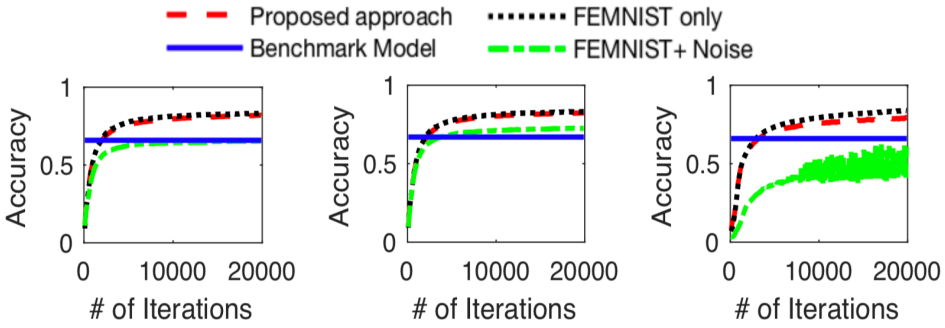
5) V is used to defined an upper limit of acceptable loss values: $\lambda^* = argmin_{\lambda} \sup_{x} |F_V(x) - F_P^{\lambda}(x)|$

6) Then, each client makes the selection of relevant data locally: $\mathcal{F}_n = \{(x_i, y_i) \in \mathcal{D}_n : l(f(x_i, \theta_{\mathcal{B}}), y_i) \leq \lambda\}$

Once the selection is made locally for every client, the standard federated learning process starts.

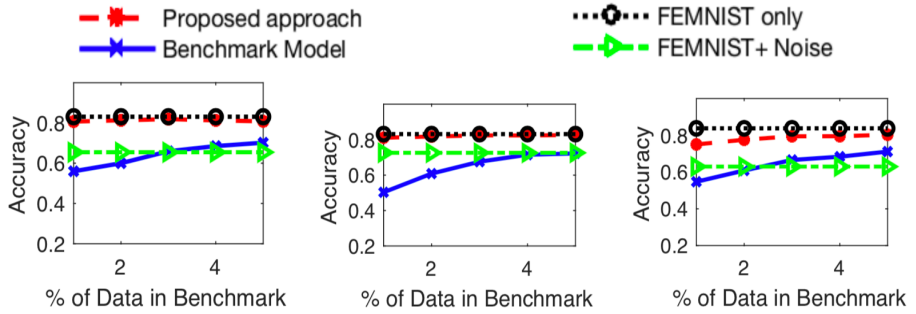


Data selection experiment 1: Different Types of Noise



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

Figure 1

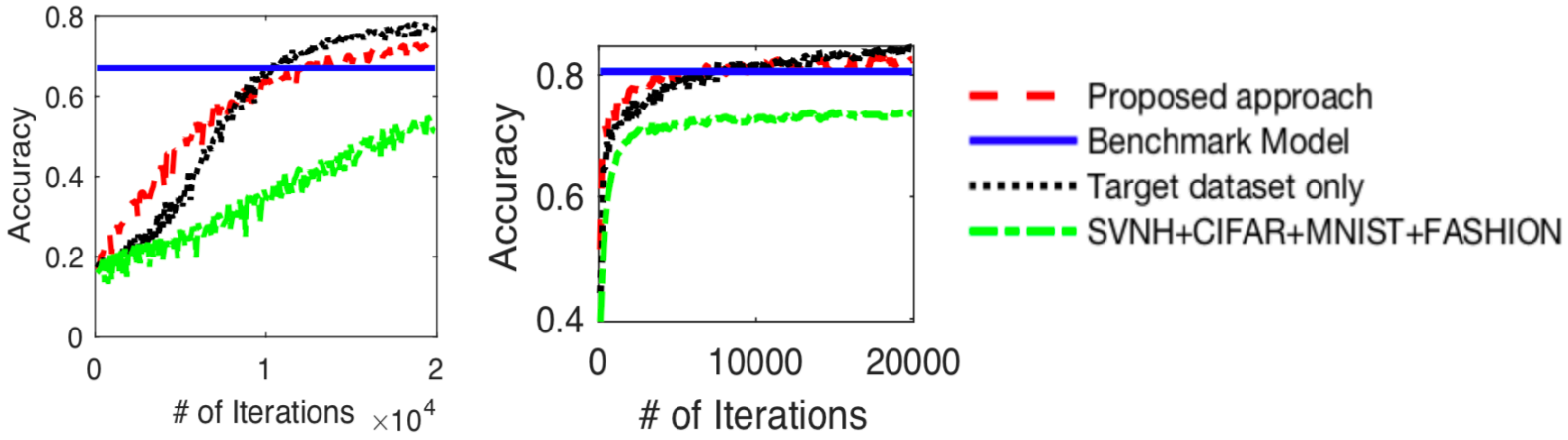


(a) Noise: CHAR-74, Target: FEMNIST (b) Noise: CIFAR-100, Target: FEMNIST (c) Noise: Closed-set, Target: FEMNIST

Figure 2

- Figure 1 shows the accuracy with different noisy data settings: open-set (Char-74, CIFAR-100) and closed-set
- Our approach always performs close to the best case line (”FEMNIST only”)
- Our approach also performs 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.
- Figure 2 shows the results of repeating the same experiments as in Figure 1 but the amount of benchmark data varied from 1% to 5% of the original dataset.
- The performance of the benchmark model increases with the benchmark dataset size while the performance of our approach remains nearly constant.

Data selection experiment 2: Strong noise



(a) Target: SVHN (b) Target: FASHION

- Our approach always performs close to the best case line even under strong noise scenario where 75% of the training data are noise