Federated Residual Learning

Chen-Yu Wei[1], Alekh Agarwal[2], John Langford[2]

[1]University of Southern California
[2]Microsoft Research

Motivation

In standard FL, only a **global model** is learned (lacking **personalization**)

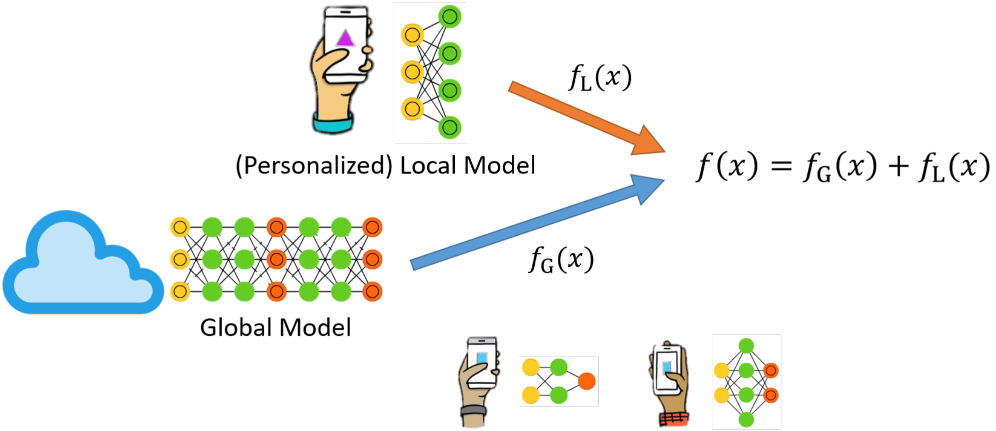
Our goal:

Building a robust FL system that

- allows personalization on top of the global model
- in the worse case, is always no worse than individual learning

(we do not assume that the clients are similar)

Proposed Framework: Federated Residual Learning



Flexibility: local models can be independently designed

FedResAvg Algorithm

For each communication round / each sampled client c

- Fetch global model $\theta_G^{(c)} \leftarrow \theta_G$
- Update local model for K times:

$$\theta_L^{(c)} \xleftarrow{-} \nabla_L \ell\left(\theta_G^{(c)}, \theta_L^{(c)}\right) \qquad \text{for } i = 1, \ldots, K$$

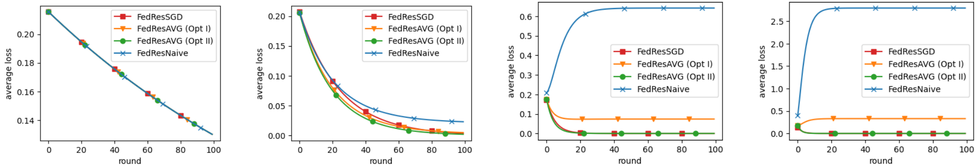
- Update (local copy of) global model for K times:

$$\theta_G^{(c)} \xleftarrow{-} \nabla_G \ell\left(\theta_G^{(c)}, \theta_L^{(c)}\right) \qquad \text{for } i = 1, \ldots, K$$

Average $\theta_G \leftarrow \frac{1}{C}\sum_c \theta_G^{(c)}$

Other technique: Using **control variates** to prevent client drift (Scaffold, ICML2020)

Experiment (Synthetic I)

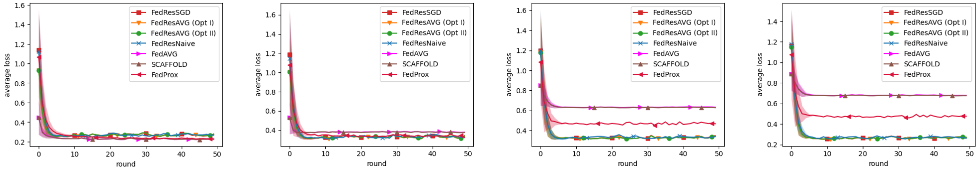


(a) $K = 1$ (b) $K = 10$ (c) $K = 50$ (d) $K = 100$

Figure: Average loss versus communication round with $N = 2$ and synthetic losses under different K (K is the number of local updates within one communication round)

$$L_1(w, \theta_1) = 0.1(w + \theta_1)^2 + 10w$$
$$L_2(w, \theta_2) = 0.1\theta_2^2 - 10w.$$

Experiment (Synthetic II)



(a) $\alpha = 0$ (b) $\alpha = 0.3$ (c) $\alpha = 0.7$ (d) $\alpha = 1$

Figure: Average loss versus communication round with $N = 10$ and synthetic losses. The loss is logistic loss on a binary classification problem whose label is generated according to Eq. (1).

Feature $x \in \mathbb{R}^d$ and label y generated by

$$y = \text{argmax}\left\{(1-\alpha)W^* x + \alpha\Theta_i^* x + n\right\} \quad (1)$$