CONVERGENCE BOUNDS FOR NODE SELECTION IN FEDERATED LEARNING

Francesco Malandrino
CNR-IEIIT, CNIT
email: francesco.malandrino@cnr.it

Carla Fabiana Chiasserini
Politecnico di Torino, Italy
CNR-IEIIT, CNIT

A priori, closed-form loss bounds have emerged as an important tool to estimate the performance of distributed learning and identifying the best learning nodes therein. Through real-world experiments, in this work we demonstrate how loss bounds are far away from the actual loss values; none the less, they can be useful for node selection purposes.

Keywords: *federated learning, loss bounds*

Federated learning (FL) [1], along with its derivative versions, is arguably the most promising distributed learning paradigm. Our work targets two of the main problems in FL, to wit:

- estimating the learning performance *before* running the learning process itself, and
- identifying the nodes that are likely to contribute the most to the learning.

We do so by performing a set of experiments leveraging the CIFAR dataset [2] and the LENET DNN [3]. We consider a total of 10 learning nodes, differing for the number of images in their local datasets (i.e., the *quantity* of data) and the classes represented therein (i.e., the *quality* of data). For the bounds, we leverage the work [4], where the expected loss at iteration t is bounded by

$$\frac{8L/\mu}{(t-1+8L/\mu)}\left(\frac{16G^2}{\mu}+4L\mathbb{E}\|\mathbf{w}_1-\mathbf{w}^*\|\right), \quad (1)$$

where

- μ is a non-negative quantity such that loss function F is μ-strongly convex;
- L is a non-negative quantity such that loss function F is L-smooth;
- G is a non-negative quantity such that the squared norm of the gradients of loss function F is bounded by G^2.

To begin with, we consider the actual and predicted loss bounds, represented in Fig. 1. Looking at the scales in the plot, it is immediately evident that the bound (1) is quite far away from the actual loss; therefore, it is of limited usefulness in directly predicting the evolution of the learning process.

We now turn our attention to the local values of the quantities appearing in (1), specifically, L, and seek to correlate them with the usefulness metric introduced in [5]. The results are summarized in Fig. 2.

It is extremely interesting to observe how, barring a single outlier, nodes with a high local value of L also have a good usefulness. These results suggest that, while bounds such as (1) may not be the best tool to quantitatively predict the evolution of learning, the quantities appearing therein – which themselves depend upon the local datasets – are an excellent tool to identify the most promising learning nodes.

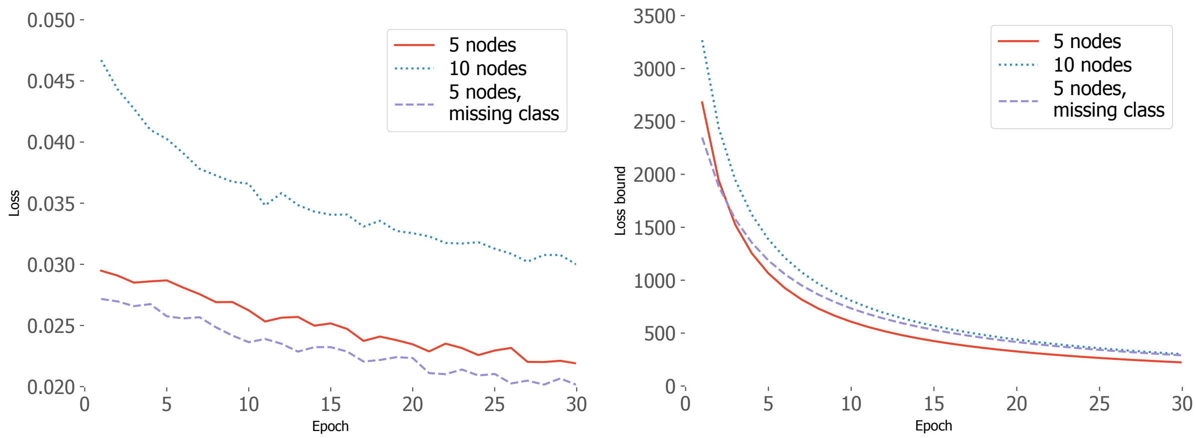

Figure 1: FL experiments: loss achieved during the training (left); bounds thereto (right).

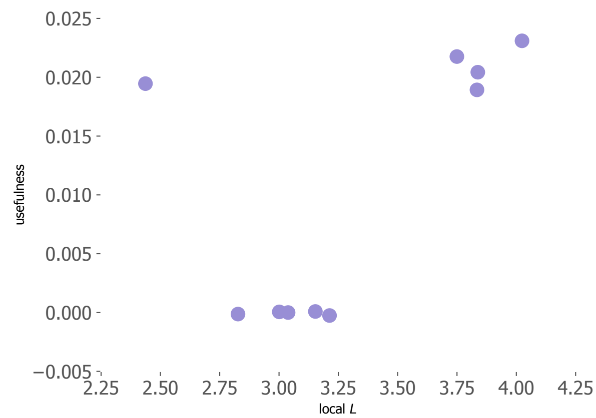

Figure 2: FL experiments: relationship between the node usefulness and the local values for the L.

REFERENCES

1. Konečný, J., McMahan, B. and Ramage, D. Federated optimization: Distributed optimization beyond the datacenter, *arXiv preprint arXiv:1511.03575*, (2015).

2. Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images, (2009).

3. Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, (1998).

4. Li, X., Huang, K., Yang, W., Wang, S. and Zhang, Z. On the convergence of fedavg on non-iid data, *International Conference on Learning Representations*, (2019).

5. Malandrino, F. and Chiasserini, C. F. Federated learning at the network edge: When not all nodes are created equal, *IEEE Communications Magazine*, (2021).