Data Resampling for Federated Learning with Non-IID Labels

Zhenheng Tang1 Zhikai Hu2 Shaohuai Shi1 Yu-ming Cheung1 Yilun Jin2 Zhenghang Ren2 Xiaowen Chu1
1Department of Computer Science, Hong Kong Baptist University
2Department of Computer Science and Engineering, The Hong Kong University of Science and Technology

Introduction

Federaled Learning performance suffers from Non-IID data [3][1]. We find that the learning process of one client (without communication) can be seen as the imbalanced learning. Intuitively, by balancing the sampling probability between labels, the label sampling probability between clients could become similar, thus the data distribution becomes label IID.

Approach

Based on the phenomenon and the findings in the preliminary results, we propose to conduct Imbalanced Weight Decay Sampling (IWDS), which decays the resampling degree with the training time, such that

1. The convergence can be accelerated at early stage.
2. The model can learn more information of its own special knowledge well in the late training stage.

From Fig. 3, we can see higher value of $\beta$ could make two different label sampling probability become more similar. So, in order to make clients have similar label sampling probability, we firstly use high $\beta$ at early training stages. And then, to diminish the effect of resampling, we gradually decay it to a value that makes $q(k_c, i_c)/q(k_c, i_c)$ close to uniform sampling case. Therefore, we change the sampling weight of $i$-th sample in client $k$ into $\alpha k c i = \frac{\lambda^k c i}{\lambda^k c i + \beta^k c i}$, in which $i$ is the communication round. The $\beta$ is updated during each communication round as

$$\beta_{t+1} = \beta_t \left(1 - \frac{1}{\alpha_{t+1}} \right).$$

Experiment settings

- Datasets and models. We evaluate our methods on CIFAR-10 with VGG-9, and Fashion-MNIST with a simple CNN used in [4].
- Datasets partition. We conduct experiments of FL with 10 clients, and 5 clients will be chosen in every communication round. For both two datasets, 4 different ways of data partition are tested: LDA partition with $w_0 = 0.5$ and $w_u = 0.1$, LLT partition with $w_0 = 0.9$ and $w_u = 0.9$.

Experiment results

- For all experiments with LLT partition, our method IWDS attain the fastest convergence rate and the highest final accuracy.
- When LDA partition, our method IWDS attain the faster convergence rate and the similar final accuracy.

References