Compression for Federated Random Reshuffling

A Preprint

Tikhon Antyshev
MIPT

Grigory Malinovsky
KAUST

Abstract

Federated Random Reshuffling (FedRR) is a recently developed method for federated training of supervised machine learning models via empirical risk minimization. It utilizes Random Reshuffling (RR), a variant of Stochastic Gradient Descent (SGD)along with Local Training carried out by the clients. We propose integration of compression techniques in FedRR, reducing the number of communicated bits in order to overcome communication bottleneck, furthermore we integrate server-side optimization (Server Stepsizes) to get improvement in theory and practice. To the best of our knowledge, this is the first time FedRR will be combined with Server Stepsizes and Compressed Iterates at the same time.

Keywords Machine Learning · Federated Learning · Random Reshuffling

1 Introduction

Modern machine learning models heavily rely on Empirical Risk Minimization for supervised training training. The success of this approach itself can be attributed to a plentiful amount of data available.

1.1 Problem Statement

We consider the standard finite-sum optimization formulation of federated learning problem:

$$x_* = \arg\min_{x\in\mathbb{R}^d} \left[f(x) \overset{\text{def}}{=} \frac{1}{M}\sum_{m=1}^{M} f_m(x) \right]$$

$$f_m(x) \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} f_m^i(x)$$

Where M is number of clients in current task, $f_m : \mathbb{R}^d \to \mathbb{R}$ is the loss of the model x on the training data subset owned by client or device $\underline{m,}$ which has the finite-sum structure, comprised of $f_m^i : \mathbb{R}^d \to \mathbb{R}$ - loss of model x on training point $i \in \overline{1, n}$. We shall also assume that $\forall\ m,\ i:\ f_m^i$ is differentiable and consider strongly convex, convex and non-convex modes.

1.2

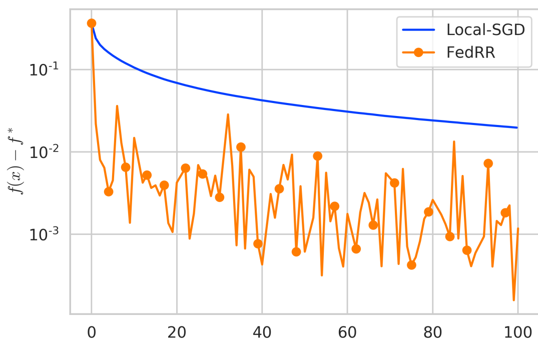
2 Contributions

3 Preliminaries

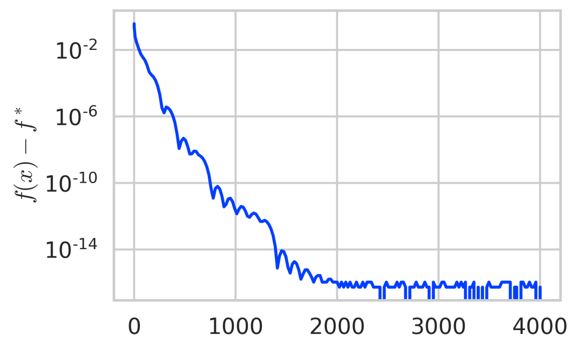
4 Theory

5 Experiments

The goal of the experiment is to demonstrate improvements of implementing Compressed Iterates with Server Step-Sizes method. We run our experiment on $l2$-regularized logistic regression with the 'a1a' dataset from LibSVM. As a baseline solution we will be demonstrating the results of Proximal and Federated RR. Afterwards, we shall compare performance for proposed algorithm and versions that use either only Server Step-sizes or only Compressed Iterates.



Local SGD and Federated RR



Performance of Nesterov method

6 Examples of citations, figures, tables, references

6.1 Citations

6.2 Figures

6.3 Tables

6.4 Lists

Список литературы