Benchmarking Federated Learning Scalability

Benchmarking Federated Learning Scalability

Gianluca Mittone[1], Samuele Fonio[1]

[1]University of Turin - Computer Science Department, Corso Svizzera, 185, Turin, Italy

Abstract

Federated Learning (FL) is a widespread Machine Learning paradigm handling distributed Big Data. In this work, we demonstrate that different FL frameworks expose different scaling performances despite adopting the same technologies, highlighting the need for a more comprehensive study on the topic.

Keywords

Federated Learning, Frameworks, Benchmark, Scalability

1. Introduction

Federated Learning (FL) [1] is a relatively recent Machine Learning (ML) paradigm focused on collaboratively training an ML model between multiple parties without sharing the local data. In particular we study the *Centralized* FL paradigm, where a central server is responsible of collecting the models from the different institutions, aggregate them without exposing local data and finally broadcast the aggregated model. Deep Neural Networks (DNNs) are naturally suited for FL tasks, since their weights are tensors and it is straightforward to aggregate them through mathematical operations. FL thus exhibits computational requirements for DNNs' training and communication requirements for exchanging them. These aspects stand out from the ML field and should be analysed from the Distributed Computing perspective.

2. Preliminary experimental results

Table 1 reports the computational performance of two mature FL frameworks, i.e., OpenFL and Flower, and one in development by the authors, i.e., FastFL. FastFL distinguishes itself for being C++-based instead of Python-based and communicating through TCP instead of gRPC. These frameworks are designed for cross-silo FL, in which a few clients (<100) possess good computational performance and stable connections. The experiments have been run on the C3S HPC centre of the University of Turin, with each client allocated on a different computational node (OmniPath network, 2 x Intel® Xeon® CPU E5-2697 v4 per node), training a ResNet18 model for 50 rounds on MNIST.

OpenFL and Flower display different scaling behaviours despite being built with the same technologies, with Flower outperforming OpenFL in both scenarios. FastFL is comparable to

PREPRINT - *The 2nd Italian Conference on Big Data and Data Science, ITADATA2023, Sept 11 – 13, Naples, IT*

gianluca.mittone@unito.it (G. Mittone); samuele.fonio@unito.it (S. Fonio)

https://alpha.di.unito.it/gianluca-mittone/ (G. Mittone); https://alpha.di.unito.it/samuele-fonio/ (S. Fonio)

0000-0002-1887-6911 (G. Mittone); 0009-0003-1870-4233 (S. Fonio)

Table 1
Execution times (s) of different FL frameworks with respect to varying numbers of clients (1 client/node)

Strong Scaling	1	2	4	8	16	32
OpenFL [2]	14967	8433	5051	4104	4870	7517
Flower [3]	14872	7672	4184	2435	1633	1415
FastFL [4]	10175	5414	2821	1656	1085	905
Weak Scaling	**1**	**2**	**4**	**8**	**16**	**32**
OpenFL	14967	15578	15853	16624	18216	—
Flower	14872	14636	14999	15046	15128	15385
FastFL	10249	9951	10090	10340	10407	10607

Flower in the weak learning case and exposes the best strong scaling properties. The first observation on FastFL is justified by the lower overhead implied by the C++ implementation, while the second by the communication backend employed: OpenFL and Flower are gRPC-based, which seems not to scale efficiently. This is evident with OpenFL 32 clients weak scaling experiment: the time is unavailable since it exceeded C3S maximum job time (6 hours).

3. Conclusions and Future Works

FL frameworks exhibit different computational performances despite being constructed with the same aim and technologies. The provided experimental results prove that it is necessary to study more in-depth the current FL frameworks from a distributed computed perspective to both investigate their real applicability scenarios and build better-performing ones.

Acknowledgments

This work receives EuroHPC-JU funding under grant no. 101034126, with support from the Horizon2020 programme (the European PILOT) and is supported by the Spoke ”FutureHPC & BigData” of the ICSC – Centro Nazionale di Ricerca in ”High Performance Computing, Big Data and Quantum Computing”, funded by European Union – NextGenerationEU.

References

[1] B. McMahan, et al., Communication-efficient learning of deep networks from decentralized data, in: AISTATS, volume 54, PMLR, 2017, pp. 1273–1282.
[2] P. Foley, et al., Openfl: the open federated learning library, Physics in Medicine & Biology 67 (2022) 214001.
[3] D. J. Beutel, et al., Flower: A friendly federated learning research framework, arXiv preprint arXiv:2007.14390 (2020).
[4] G. Mittone, et al., Experimenting with emerging risc-v systems for decentralised machine learning, in: CF ’23, ACM, 2023.