Towards Efficient and Practical Federated Learning

Ahmed M. Abdelmoniem
Queen Mary University of London, UK

Abstract

Federated learning (FL) is increasingly becoming the norm for training models over distributed and private datasets. In this poster, we describe our efforts which are motivated by the urging need for exploring the promising prospects and imminent challenges towards the facilitation of the adoption of federated learning for service providers. The prospects are motivated by the growing interest and momentum towards the adoption of privacy preservation and 5G/6G technologies. Whereas, the challenges that hinder wide FL adoption are mainly the data, resource and user heterogeneity as well as communication overhead. Resource heterogeneity is challenging in particular because the training resources and conditions are not under the control of the central server. Specifically, FL-trained models in practice require a significant amount of time (days or even weeks) because FL tasks execute in highly heterogeneous environments where devices only have widespread yet limited computing capabilities and network connectivity conditions.

Introduction

We present our initial efforts focused on studying training FL models in heterogeneous environments. We note that heterogeneity can have a detrimental impact on both the model quality and fairness. Then, we present our efforts towards designing efficient yet practical mitigation methods to limit the impact of system heterogeneity in FL settings.

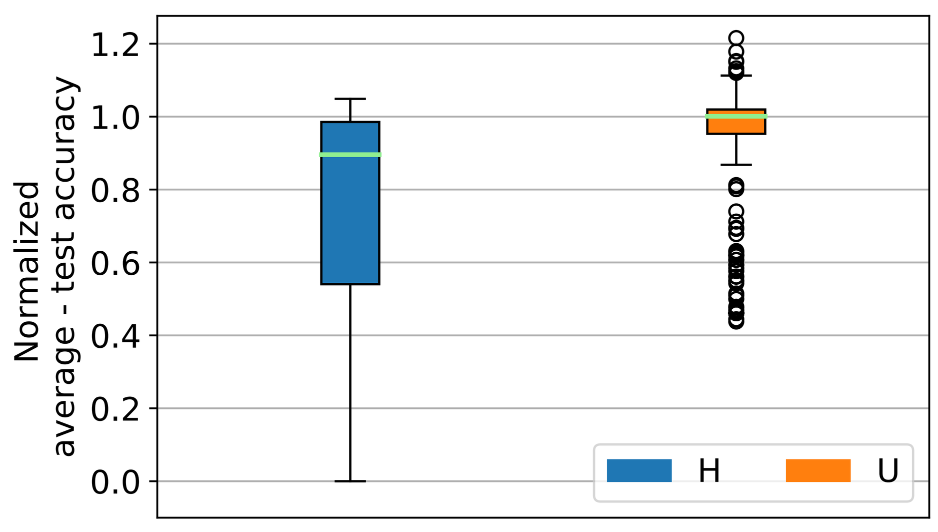
Federated Learning in A Nutshell

Federated Learning (FL) is a recent paradigm that has sparked great research interest to enable learning collaboratively over distributed datasets [1, 2, 3]. In FL, with the help of a central aggregation server, the end devices train a global model on their private data without transferring the private data over the network. Even though, the paradigm was initially proposed by the industry (e.g., [4, 3]) to solve a practical problem, it has seen tremendous interest from academia (e.g., [5, 6]).

Empirical Study of Heterogeneity [7]

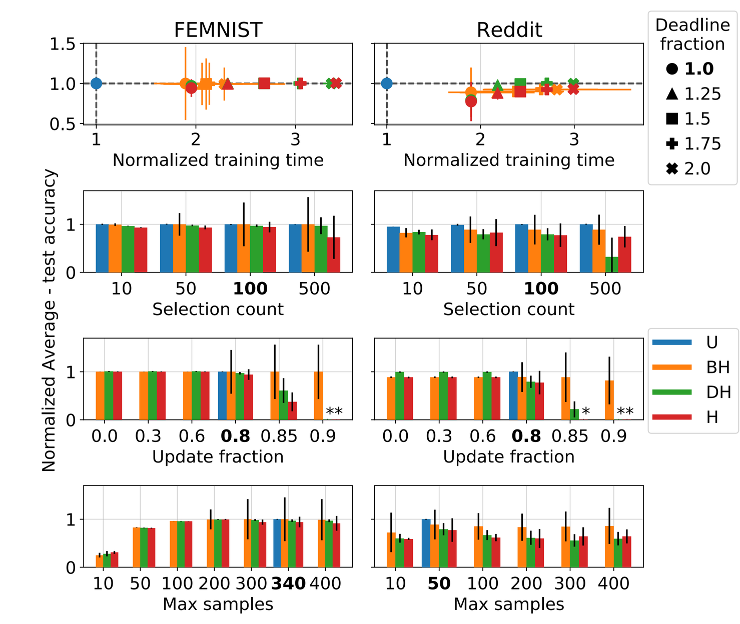
Fig. 1: contrasts two scenarios: (**U**) the ideal case where devices have uniform system capabilities and are sampled uniformly; and (**H**) the realistic case with heterogeneous hardware/network and variable availability over time. The test accuracy is averaged across 1.5K experimental configurations for five FL benchmarks (e.g., natural language processing and image classification). The average test accuracy in **H**, on median is below 0.88×, and at the extreme, the training diverges.
Fig. 2: shows the sensitivity of the average accuracy to the FL hyper-parameters' choice in FEMNIST and Reddit Benchmarks. We observe that the average performance is quite sensitive to the choice of the FL hyper-parameters such as the deadline, selection count, update fraction, and max samples. The choice of these hyper-parameters influences whether devices and how many are able to meet the reporting deadline.

Fig. 1: Model Quality with System Heterogeneity



H heterogeneous cases — **U** ideal cases

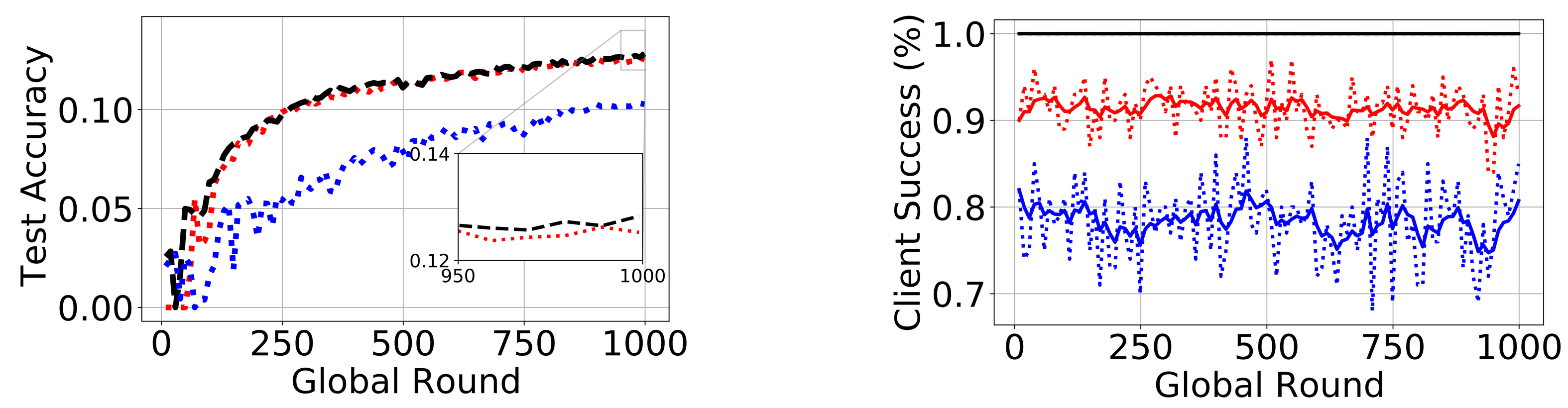
Fig. 2: Influence of Hyper-parameter Selection



AQFL: Adaptive Model Quantization [8]

AQFL is an adaptive model quantization technique that varies the quantization level in proportion to the resources available at clients' end-devices. This approach reduces both the computational and communication overhead of the training process, which helps with increasing the percentage of devices submitting their model updates within the deadline.
Fig. 3: shows, for Reddit benchmark, that **H** (blue line) degrades the model quality significantly by 25% and **AQFL** (red line) can mitigate the impact from heterogeneity by maintaining the model quality of the ideal case (black line). AQFL can customize the quantization level per devices' capabilities, which increases the success rate by nearly 12%.

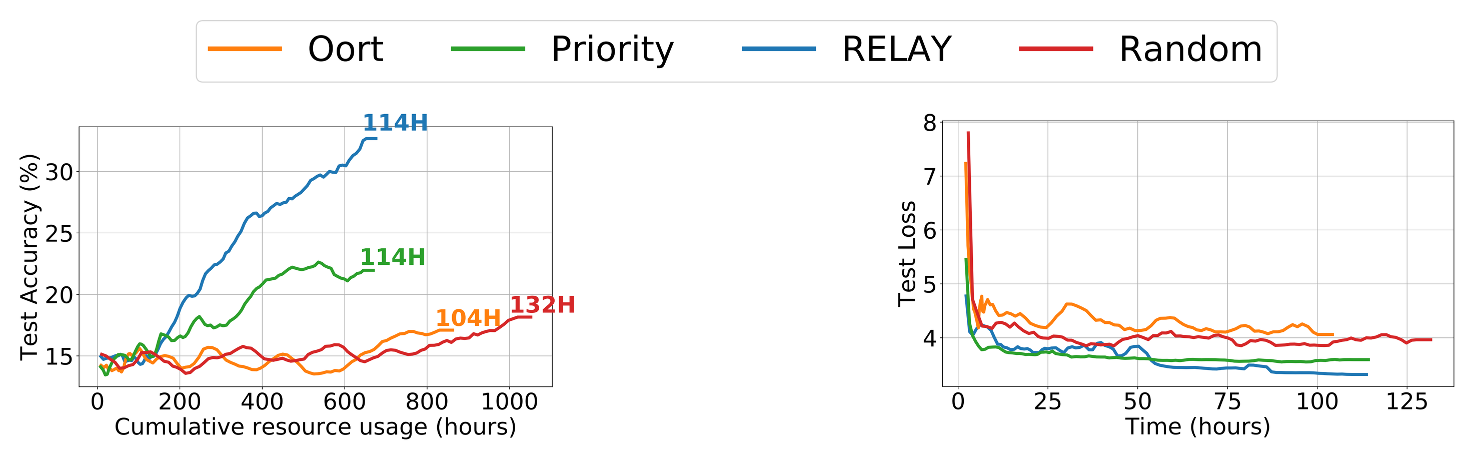
Fig. 3: Performance with vs without AQFL



RELAY: Resource Efficient Federated Learning [9]

RELAY maximizes existing FL systems' statistical, system, and resource efficiency with little modification to the existing FL solutions. It decouples the collection of participant updates from aggregation into an updated model. It also intelligently selects among available participants that are least likely to be available in the future.
Fig. 4: shows the performance of RELAY vs selection methods including SOTA system (Oort [10]). The average test accuracy vs resource usage is on left and test loss vs time is on right. RELAY achieves superior performance over other methods because it maximizes resource diversity and reduces wastage thanks to availability selection and stale aggregation.

Fig. 4: RELAY vs State-of-the-Art Systems



Summary

Our main findings towards studying and mitigating the effects of heterogeneity are:

1. Our empirical study shows that system heterogeneity can result in degradation of 5X on average for model quality [7].
2. AQFL is an adaptive model quantization method which can homogenize the device training time and maintain the quality and fairness [8].
3. RELAY is a system enabling staleness-aware aggregation and intelligent participant selection to improve resource efficiency and model quality [9]

References

[1] Jakub Konečný et al. "Federated Learning: Strategies for Improving Communication Efficiency". In: *Workshop on Private Multi-Party Machine Learning - NeurIPS*. 2016.
[2] H. Brendan McMahan et al. "Communication-Efficient Learning of Deep Networks from Decentralized Data". In: *AISTATS*. 2017.
[3] Keith Bonawitz et al. "Towards Federated Learning at Scale: System Design". In: *MLSys*. 2019.
[4] Timothy Yang et al. "Applied Federated Learning: Improving Google Keyboard Query Suggestions". In: *arXiv 1812.02903* (2018).
[5] Sebastian Caldas et al. "LEAF: A Benchmark for Federated Settings". In: *arXiv 1812.01097* (2018).
[6] Chengxu Yang et al. "Characterizing Impacts of Heterogeneity in Federated Learning upon Large-Scale Smartphone Data". In: *The Web*. 2021.
[7] Ahmed M. Abdelmoniem et al. "Empirical Analysis of Federated Learning in Heterogeneous Environments". In: *ACM EuroMLSys*. 2022.
[8] Ahmed M. Abdelmoniem and Marco Canini. "Towards Mitigating Device Heterogeneity in Federated Learning via Adaptive Model Quantization". In: *ACM EuroMLSys*. 2021.
[9] Ahmed M. Abdelmoniem et al. "Resource-Efficient Federated Learning". In: *arXiv 2111.01108* (2021).
[10] Fan Lai et al. "Efficient Federated Learning via Guided Participant Selection". In: *USENIX OSDI*. 2021.