Collaborative Learning Experimentation Testbed

Janez Božič*
University of Ljubljana

Amandio R. Faustino*
KAUST

Boris Radovič*
KAUST, University of Ljubljana

Marco Canini
KAUST

Veljko Pejović
University of Ljubljana

Progressing beyond centralized AI is of paramount importance, yet it has been shown that distributed AI solutions, in particular various federated learning (FL) algorithms, cannot be realistically assessed in a simulation setting [6], which prevents the research community from identifying the most promising approaches and practitioners from being convinced that a solution is deployment-ready. The largest hurdle towards FL algorithm evaluation is the difficulty of conducting real-world experiments over a variety of FL client devices and different platforms, while evaluating various dimensions of algorithm performance, such as inference accuracy, energy consumption, and time to convergence.

In this paper, we overcome the limitations of previous work that either supported experimentation with a single algorithm and unified hardware architectures [6] or only used trace-based simulation [3], and present CoLExT, a real-world testbed for FL research. CoLExT is designed to streamline FL experimentation over heterogeneous edge devices, such as single-board computers (SBCs) and smartphones, while providing real-time collection and visualization of performance metrics. CoLExT only requires minimal developer's effort to port algorithms to testbed-based execution, and its metrics collection instrumentation introduces negligible resource usage overhead. Furthermore, CoLExT includes an easy-to-use configuration mechanism allowing an experimenter to modify the execution scenario, including the number/nature of devices running the algorithm, the training arguments, and metric collection settings. Finally, through our initial experimentation, we reveal previously unknown trade-offs, inefficiencies, and programming bugs that popular FL algorithms exhibit once evaluated in a realistic setting and, thus, demonstrate the practical usability of CoLExT.

Figure 1 illustrates how CoLExT sheds new light on FL algorithm evaluation. We experiment with 3 FL methods (FedAvg [5], FedProx [4], and HeteroFL [2]), with different model sizes and a different ratio of clients participating per training round. Traditional means of simulation-based assessment would only compare algorithms based on their accuracy and would identify FedProx with a large model and full client participation as the most promising solution

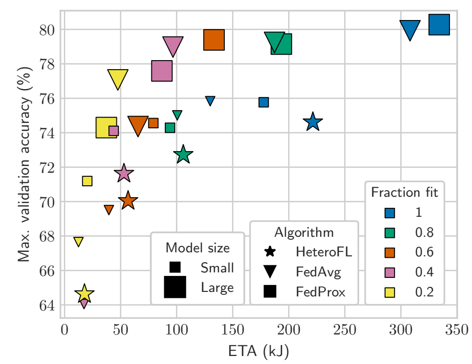

Figure 1: Validation accuracy VS energy to accuracy.

(top-right blue square). CoLExT, on the other hand, uncovers other metrics, such as CPU utilization, training duration, and energy usage. In the figure, CoLExT reveals that the amount of energy needed for reaching a particular level of accuracy (i.e., energy-to-accuracy, ETA) differs drastically among points that achieve very similar accuracy. Thus, while the previously identified FedProx configuration reaches the top accuracy, it consumes 3× more energy for a mere 4% increase in accuracy when compared to the FedAvg configuration using 40% client participation (top-left pink triangle).

For more details on CoLExT, see our technical report [1].

References

[1] Janez Božič, Amândio R. Faustino, Boris Radovič, Marco Canini, and Veljko Pejović. 2024. *Where is the Testbed for my Federated Learning Research?* https://arxiv.org/abs/2407.14154

[2] Enmao Diao, Jie Ding, and Vahid Tarokh. 2021. HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients. In *ICLR*.

[3] Fan Lai, Yinwei Dai, Sanjay S. Singapuram, Jiachen Liu, Xiangfeng Zhu, Harsha V. Madhyastha, and Mosharaf Chowdhury. 2022. FedScale: Benchmarking Model and System Performance of Federated Learning at Scale. In *ICML*.

[4] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2020. Federated Optimization in Heterogeneous Networks. In *MLSys*.

[5] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *AISTATS*.

[6] Kok-Seng Wong, Manh Nguyen-Duc, Khiem Le-Huy, Long Ho-Tuan, Cuong Do-Danh, and Danh Le-Phuoc. 2023. An Empirical Study of Federated Learning on IoT-Edge Devices: Resource Allocation and Heterogeneity. arXiv:2305.19831 [cs.LG]

*Equal contribution. Work partly done while Janez Božič interned at KAUST.