Poster: HARMONY: Heterogeneity-Aware Hierarchical Management for Federated Learning System

Chunlin Tian [1] **Li Li** [1] **Zhan Shi** [2] **Jun Wang** [3] **ChengZhong Xu** [1]

1 Motivation

Federated learning (FL) enables multiple devices to collaboratively train a shared model while preserving data privacy. However, despite its emerging applications in many areas, real-world deployment of FL across heterogeneous edge devices is challenging due to devices having wildly diverse training capability and data distribution, which highly impact both model performance and training efficiency.

We discuss the key observations around system and data heterogeneity. Specifically, we try to derive critical design principles for an effective FL system:

#Principle 1: Accurate estimation of the runtime device training capability is critical to the system efficiency of an FL system.

#Principle 2: Homogeneous local training data distributions are beneficial to the statistical efficiency of the global model.

#Principle 3: In addition to the local data distribution, a homogeneous overall training dataset is also critical to the statistical efficiency of the global model.

According to the above discussion, we can find that to intelligently strike a balance between the training efficiency and the model performance, Harmony needs to jointly consider the runtime training capability from the system perspective and both the local and overall data distribution from the data perspective.

2 Limitations of the State of the Art

To improve the model performance or training efficiency, several device selection approaches (Li et al., 2021; Nishio & Yonetani, 2019; Chai et al., 2020) have been proposed. However, most of them decouple the problem and only focus on one certain side. For instance, AutoFL (Kim & Wu, 2021) uses a reinforcement learning based approach for device selection in order to jointly optimize the FL training time and energy efficiency of individual participants. However, only focusing on the system heterogeneity can unconsciously select devices with high data skewness in the training process which severely hampers model performance and limits the application scope in real-world scenarios. Though the system in (Lai et al., 2021) conducts device selection by jointly considering the system configuration and the training loss of a specific device, this coarse-grained selection ignores the skewness of the overall data distribution for each training round which not only causes unnecessary local training, but also negatively impacts the convergence of the collaborative model.

3 Key Insights

This paper proposes Harmony[1], a high-performance FL framework with heterogeneity-aware hierarchical management of training devices and training data. Unlike previous works that mainly focus on heterogeneity in either training capability or data distribution, Harmony adopts a hierarchical structure to jointly handle both heterogeneities in a unified manner.

It effectively directs the training process to make it proceed in harmony through intelligently mediating the conflict caused by the heterogeneity in the following four folds: 1) the static system heterogeneity caused by different hardware configurations, 2) the dynamic system heterogeneity caused by resource contention at runtime, 3) the data heterogeneity in each local device and 4) the data heterogeneity in each global training round.

4 Main Artifacts

Figure 1 shows the system architecture of Harmony that mainly contains the following two components: the global coordinator hosted by the central server and the local coordinator deployed on each participating mobile device. Without accessing the raw data, the global coordinator first selects

*Equal contribution [1]University of Macau [2]The University of Texas at Austin [3]Futurewei Technology. Correspondence to: Chunlin <tianclin0212@gmail.com>, Li Li <LLiLi@um.edu.mo>.

Proceedings of the 5^{th} MLSys Conference, Santa Clara, CA, USA, 2022.

[1]The name of the system, Harmony, has two-fold implications. First, our FL system aims to manage all heterogeneous devices to work in harmony. Second, our FL system aims to have the background training task execute harmoniously with the foreground applications of the devices.

the participants and then further reorganizes the training samples within them based on the accurate estimation of the runtime training capability and data distribution of each device. The local coordinator keeps monitoring the local training status and conducts efficient training with guidance from the global coordinator.

The system workflow of Harmony can be represented as the following main steps. 1. At the initialization step, all the mobile devices participate in the first training round and complete local training. 2. The local coordinator sends the following information to the central server including: a) static system info (CPU frequency in this case), b) dynamic system info (resource contention caused by concurrently running apps), c) size of local training data, and d) the local training gradients. 3. With the received information, the global coordinator well estimates the data distribution and predicts the runtime training capability. 4. After that, the global coordinator intelligently selects the participating devices by jointly considering the homogeneity of the local training data and runtime training capability. Moreover, to strike a more effective balance, the global coordinator fine-tunes the distribution of the overall training data in the current training round to make it more homogeneous. 5. The global coordinator then broadcasts the coordination result to the corresponding selected devices. 6. The local coordinator then conducts the local training process based on the coordination result. 7. Meanwhile, the local coordinator will monitor real-time status and data. If the variance exceeds a predefined threshold, it triggers the coordination procedure. Otherwise, the training process proceeds normally without introducing extra computing and communication overhead.

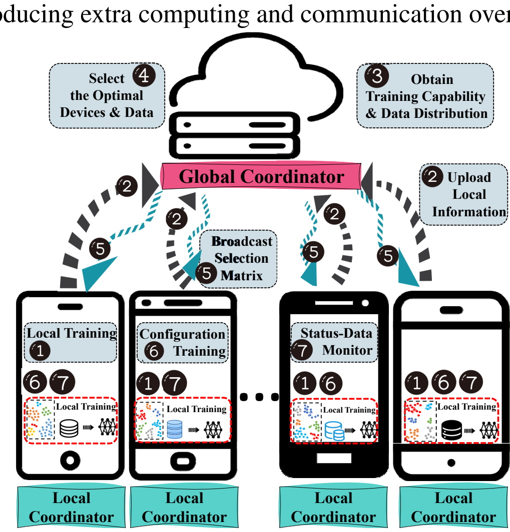

Figure 1. The system overview of Harmony.

5 Key Results and Contributions

We conduct extensive experiments to evaluate the effectiveness of Harmony. Compared with the baselines, the evaluation experiments demonstrate that Harmony improves the accuracy of the trained model by up to 27.62%, speeds up the overall training time by up to $3.29\times$, and saves up to 88.41% in energy. To our best knowledge, Harmony is the *first* work that studies the balance between the model performance and training progress through intelligently considering the static and dynamic heterogeneity from the system perspective as well as harmonizing the data heterogeneity from both the local and global levels. Specifically, we make the following key contributions:

- We propose Harmony, a high-performance federated learning framework, where the global and local coordinators cooperate to intelligently balance the model performance and training process in the highly dynamic and heterogeneous training environment.
- We design the hierarchical coordination mechanism that well coordinates the training process that jointly considers the data distribution of the local device and the overall distribution of each training round in order to effectively improve the model accuracy while accelerating the training progress.

References

Chai, Z., Ali, A., Zawad, S., Truex, S., Anwar, A., Baracaldo, N., Zhou, Y., Ludwig, H., Yan, F., and Cheng, Y. Tifl: A tier-based federated learning system. In *Proceedings of the 29th International Symposium on High-Performance Parallel and Distributed Computing*, pp. 125–136, New York, 2020. ACM.

Kim, Y. G. and Wu, C.-J. Autofl: Enabling heterogeneity-aware energy efficient federated learning. In *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 183–198, Athens, 2021. IEEE/ACM.

Lai, F., Zhu, X., Madhyastha, H. V., and Chowdhury, M. Oort: Efficient federated learning via guided participant selection. In *15th USENIX Symposium on Operating Systems Design and Implementation (OSDI 21)*, pp. 19–35, California, July 2021. USENIX Association. ISBN 978-1-939133-22-9. URL `https://www.usenix.org/conference/osdi21/presentation/lai`.

Li, Q., He, B., and Song, D. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10713–10722, virtual, 2021. IEEE.

Nishio, T. and Yonetani, R. Client selection for federated learning with heterogeneous resources in mobile edge. In *ICC 2019-2019 IEEE international conference on communications (ICC)*, pp. 1–7, Shanghai, 2019. IEEE.