

Advancing Federated Learning by Addressing Data and System Heterogeneity

Yiran Chen

Department of Electrical and Computer Engineering, Duke University
yiran.chen@duke.edu

Abstract

In the emerging field of federated learning (FL), the challenge of heterogeneity, both in data and systems, presents significant obstacles to efficient and effective model training. This talk focuses on the latest advancements and solutions addressing these challenges.

The first part of the talk delves into data heterogeneity, a core issue in FL, where data distributions across different clients vary widely and affect FL convergence. We will introduce the **FedCor (Tang et al. 2022)** framework addressing this by modeling loss correlations between clients using Gaussian Process and reducing expected global loss. External covariate shift in FL is uncovered, demonstrating that normalization layers are crucial, and layer normalization proves effective. Additionally, class imbalance in FL degrades performance, but our proposed **Fed-CBS (Zhang et al. 2023)** mechanism reduces this imbalance by employing homomorphic encryption for privacy preservation.

The second part of the talk shifts focus to system heterogeneity, an equally critical challenge in FL. System heterogeneity involves the varying computational capabilities, network speeds, and other resource-related constraints of participating devices in FL. To address this, we introduce **FedSEA (Sun et al. 2022)**, which is a semi-asynchronous FL framework that addresses accuracy drops by balancing aggregation frequency and predicting local update arrival. Additionally, we discuss **FedRepre (Wang et al. 2024)**, a framework specifically designed to enhance FL in real-world environments by addressing challenges including unbalanced local dataset distributions, uneven computational capabilities, and fluctuating network speeds. By introducing a client selection mechanism and a specialized server architecture, FedRepre notably improves the efficiency, scalability, and performance of FL systems.

Our talk aims to provide a comprehensive overview of the current research and advancements in tackling both data and system heterogeneity in federated learning. We hope to highlight the path forward for FL, underlining its potential in diverse real-world applications while maintaining data privacy and optimizing resource usage.

References

Sun, J.; Li, A.; Duan, L.; Alam, S.; Deng, X.; Guo, X.; Wang, H.; Gorlatova, M.; Zhang, M.; Li, H.; et al.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

2022. FedSEA: A Semi-Asynchronous Federated Learning Framework for Extremely Heterogeneous Devices. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, 106–119.

Tang, M.; Ning, X.; Wang, Y.; Sun, J.; Wang, Y.; Li, H.; and Chen, Y. 2022. FedCor: Correlation-based active client selection strategy for heterogeneous federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10102–10111.

Wang, Y.; Tang, M.; Chen, H.; Li, S.; Zheng, Q.; Chang, A.; Hao, C.; Li, H.; and Chen, Y. 2024. An Efficient and Scalable Federated Learning Framework with Client Representative Mechanism and Specialized Server Architecture. *preprint*.

Zhang, J.; Li, A.; Tang, M.; Sun, J.; Chen, X.; Zhang, F.; Chen, C.; Chen, Y.; and Li, H. 2023. Fed-cbs: A heterogeneity-aware client sampling mechanism for federated learning via class-imbalance reduction. In *International Conference on Machine Learning*, 41354–41381. PMLR.