

Fairness-Aware Resource Optimization in Federated Learning

Federated learning (FL) [1] is a decentralized machine learning (ML) paradigm in which multiple devices collaboratively train a shared model without exposing their private data. Unlike traditional ML techniques, FL preserves data privacy and alleviates computational load on the server by distributing the training process across devices. These advantages make FL particularly appealing for privacy-sensitive domains such as healthcare and finance. However, deploying FL in practice introduces significant challenges. In particular, devices often face limited and fluctuating resources, coupled with heterogeneity in hardware capabilities and data distributions. As a result, devices contribute unevenly in terms of data quantity, quality, and computational power. This imbalance raises a critical question: how can we ensure fair credit assignment while still optimizing communication and computation efficiency? Without mechanisms to address fairness, devices with limited resources may lose motivation to participate, ultimately threatening the scalability and sustainability of FL.

Recent work on contribution valuation in vertical FL, such as VerFedSV [2], has explored approaches based on Shapley value to quantify contributions and ensure fairness across parties holding different features. Other lines of research have examined data valuation methods, including influence functions [3] and gradient-based metrics, to capture the marginal impact of individual contributions. However, most of these efforts focus on vertical FL or simplified theoretical settings, where computational and communication costs are not the primary bottlenecks. In horizontal FL, the problem of fair contribution valuation remains largely unresolved. The challenge becomes even more pressing when fairness must be balanced against practical constraints such as communication latency, energy efficiency, and system heterogeneity.

In this work, we propose a fairness-aware resource optimization framework for horizontal FL. Instead of relying on computationally prohibitive Shapley-value calculations, our approach approximates contribution scores by evaluating each device’s marginal impact on global model improvement. In particular, we employ gradient alignment to capture how well local updates align with the global optimization direction. Furthermore, we use the update magnitude to quantify the contribution. These scores are then integrated into the resource allocation process: devices with higher contributions are prioritized in client selection, allocated more communication bandwidth, and assigned more favorable compression levels. This joint design of fairness and efficiency ensures that resource optimization not only minimizes latency and energy but also preserves equitable participation among heterogeneous devices. Ultimately, it enables FL systems that are both scalable and robust in real-world deployments.

We assess the potential of our framework, showing that it can achieve faster convergence and reduced communication costs compared to conventional resource allocation approaches, while maintaining a fair balance of device participation. This approach highlights a new direction for resource allocation in FL, where fairness and efficiency reinforce each other, paving the way for scalable and practical FL deployments.

References

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In AISTATS.
- [2] Z. Fan, H. Fang, Z. Zhou, J. Pei, M. P Friedlander, and Y. Zhang. 2024. Fair and efficient contribution valuation for vertical federated learning. In ICLR.
- [3] K. P. Wei, and L. Percy. 2017. Understanding black-box predictions via influence functions. In ICML.