

Disentangling Data Distribution for Optimal and Communication-Efficient Federated Learning (Abstract Reprint)

Xinyuan Zhao¹, Hanlin Gu², Lixin Fan², Yuxing Han¹ and Qiang Yang^{2,3}

¹ Shenzhen International Graduate School, THU

² AI Lab, Webank

³ Department of Computer Science and Engineering, HKUST

Abstract Reprint. This is an abstract reprint of a journal article by [Zhao *et al.*, 2026].

Abstract

Federated Learning (FL) facilitates collaborative training of a global model whose performance is boosted by private data owned by distributed clients, without compromising data privacy. Yet the wide applicability of FL is hindered by entanglement of data distributions across different clients. This paper demonstrates for the first time that by disentangling data distributions FL can in principle achieve efficiencies comparable to those of distributed systems, requiring only one round of communication. To this end, we propose a novel FedDistr algorithm, which employs stable diffusion models to decouple and recover data distributions. Empirical results on the CIFAR100 and DomainNet datasets show that FedDistr significantly enhances model utility and efficiency in both disentangled and near-disentangled scenarios while ensuring privacy, outperforming traditional federated learning methods.

References

[Zhao *et al.*, 2026] Xinyuan Zhao, Hanlin Gu, Lixin Fan, Yuxing Han, and Qiang Yang. Disentangling data distribution for optimal and communication-efficient federated learning. *Artificial Intelligence*, 351:104455, 2026.