

Machine Learning from Imbalanced Data Sources

Shiqiang Wang
IBM Research

Thursday, September 26, 2024
3:00 P.M. • MSEE 112

Zoom: <https://purdue-edu.zoom.us/j/95619760242> ~ **Meeting ID:** 956 1976 0242

Abstract

Data is essential for machine learning, but in real-world scenarios, it is often distributed unevenly across sources with varying sizes, quality, and availability. This imbalance poses challenges to model training, potentially leading to biased outcomes. This talk addresses these challenges through the lens of federated learning (FL), a prominent use case where each client acts as an individual data source contributing to a global model. In FL, clients may participate in training at different frequencies that are unknown a priori, meaning that data from less active clients has less impact on the model, which can create further imbalance and bias. Building on our recent work, I will introduce a lightweight algorithm, FedAU, that dynamically adapts aggregation weights in the federated averaging (FedAvg) process to handle unknown and varying client participation rates effectively. By estimating optimal weights based on the participation history of clients, FedAU ensures fair representation of all data sources, mitigating biases and enhancing convergence. I will present the key ideas of our theoretical convergence analysis that connects estimation error with convergence guarantees, as well as experimental results validating the effectiveness of FedAU under various participation patterns. Afterwards, I will discuss how the principles and methodologies developed in this work can be extended beyond FL to address the general problem of learning from imbalanced data sources across various machine learning paradigms, providing new insights and tools for enhancing model performance and fairness in diverse settings.

Bio

Shiqiang Wang is a Staff Research Scientist at IBM T. J. Watson Research Center, NY, USA. He received his Ph.D. from Imperial College London, United Kingdom, in 2015. His current research focuses on the intersection of distributed computing, machine learning, networking, and optimization, with a broad range of applications including data analytics, edge-based artificial intelligence (Edge AI), Internet of Things (IoT), and future wireless systems. He received the IEEE Communications Society (ComSoc) Leonard G. Abraham Prize in 2021, IEEE ComSoc Best Young Professional Award in Industry in 2021, IBM Outstanding Technical Achievement Awards (OTAA) in 2019, 2021, 2022, and 2023, and multiple Invention Achievement Awards from IBM since 2016. For more details, please visit his homepage at: <https://shiqiang.wang>

Host

Chris Brinton, cgb@purdue.edu, 765-494-3538