Federated Learning: Evaluating Client Roles in Decentralized Networks 🔗

Abstract

In the healthcare domain, data privacy is paramount. This project explores Federated Learning (FL) as a solution to train deep learning models across decentralized medical institutions without sharing sensitive patient data. Using a dataset of COVID-19 chest X-ray images, we simulate a federated setup with three clients and evaluate how individual client participation impacts the global model's performance under seven different data distribution scenarios. Our focus lies in quantifying each client's contribution and benefit to the central network under realistic, heterogeneous conditions.

Research Objectives

- To simulate real-world data heterogeneity using scenarios like label skew, concept shift, and quantity skew.
- To implement and train various CNN architectures (TinyCNN, SimpleCNN, ResNet18, MobileNetV2) across three clients in a federated setting.
- To evaluate both **client contribution** (using Leave-One-Out accuracy drops) and **client benefit** (accuracy gain through collaboration).
- To analyze how model complexity and data distribution affect fairness and utility in FL environments.
- To offer insights for equitable model training across healthcare institutions in a privacy-preserving manner.

Methodology

The system is implemented using Flower (FLwr) and PyTorch on Google Colab with A100 GPU support. The dataset comprises COVID-19 chest X-ray images, preprocessed and distributed among three clients. Each client trains a CNN model locally, and the server aggregates weights to update the global model over multiple rounds. We designed seven data partitioning scenarios to reflect real-world heterogeneity. Contribution is measured by excluding each client during training and observing global accuracy drop, while benefit is measured by comparing a client’s isolated accuracy with its federated performance. All results and metrics are logged per round and stored for analysis.

Expected Outcomes

- Identification of patterns where certain clients consistently contribute more to model performance.
- Insights into scenarios where data diversity improves global generalization.
- Visual dashboards (via Streamlit) to compare contribution and benefit metrics across models and distributions.
- Guidelines for fair client selection or incentivization strategies in healthcare FL deployments.

Conclusion

This project demonstrates the feasibility and importance of evaluating client roles in federated learning, particularly in sensitive sectors like healthcare. By quantifying both contribution and benefit under diverse conditions, it paves the way for more transparent, equitable, and efficient FL systems. The findings aim to guide practitioners in designing better collaboration protocols that balance performance with privacy.