

Fairness-Aware Federated Learning for Text Classification

Shubham Gupta, Manas Jorvekar

Department of Computer Science & Engineering, COEP Technological University, Pune

Guide: Dr. Y.V. Haribhakta

Department of Computer Science & Engineering, COEP Technological University, Pune

Field Targeted:

Artificial Intelligence, Machine Learning, and Natural Language Processing (NLP), privacy-preserving federated learning for text classification.

Future Prospects & Importance:

- Federated Learning (FL) is crucial for privacy in AI, enabling collaborative model training across decentralized devices without sharing raw data.
- Applications span healthcare, IoT, e-commerce, and secure NLP solutions, addressing real-world needs for data privacy and fairness.
- Ensuring fairness and robust performance in non-IID (non-Independent and Identically Distributed) data settings is vital for widespread, ethical adoption of FL in industry and research.

Project Abstract:

We propose *DualMetric-Adaptive FL*, a novel federated learning algorithm designed to enhance both global performance and client-level fairness in multi-class text classification tasks, particularly under challenging non-IID data distributions. Our method introduces dual-metric monitoring and adaptive regularization to ensure robust, fair, and privacy-preserving NLP solutions for real-world federated environments.

Novelty:

- Developed a DualMetric-Adaptive FL algorithm that:
 - Monitors both global and local (client-specific) accuracies to measure overall performance and fairness.
 - Dynamically adjusts the regularization parameter for each client based on intra-client and inter-client performance trends, unlike traditional static methods.
- Demonstrated significant improvements in both global accuracy and fairness (lower variance among clients) compared to existing methods like FedProx.

How Will It Be Used?

- Enables organizations to collaboratively train NLP models (e.g., for chatbots, sentiment analysis) without compromising user privacy.
- Particularly beneficial in settings where data is distributed unevenly (non-IID), such as across different branches or user devices.
- Can be integrated into federated learning frameworks (e.g., Flower) for scalable deployment in real-world applications.

Key Results:

Metric	FedProx (Baseline)	DualMetric-Adaptive FL (Proposed)	Improvement
Global Accuracy (500 rounds)	87.14%	92.39%	+5.25%
Fairness (Variance, 500 rounds)	0.0055	0.0006	-89.09%

- Achieved higher and more stable global accuracy.
- Substantially reduced bias and variance among client performances.

Specific Details:

- Dataset: Bitext retail e-commerce chatbot dataset (45k samples, 46 categories).
- Simulation: 11 clients with non-IID data splits, evaluated over 500 training rounds.
- Framework: Implemented using Flower for federated learning experimentation.

Conclusion:

Our DualMetric-Adaptive FL framework advances federated text classification by dynamically balancing accuracy and fairness, minimizing bias, and maintaining privacy - setting a new benchmark for federated NLP in real-world, heterogeneous environments.