Chenguang Xiao (He/Him)

 c.g.xiao@outlook.com | +44 7421731656

 xiao-chenguang.github.io | github.com/Xiao-Chenguang

Summary: PhD in Computer Science with 5+ years of research experience in deep learning and AI, specializing in federated learning, model optimization, and large-scale machine learning systems. Proficient in developing and evaluating advanced AI models, with strong mathematical foundations and programming expertise. Seeking to leverage my skills to contribute to ground-breaking AI initiatives

EDUCATION

PhD in Computer Science, University of Birmingham, UK   2020-2024
- <u>Research areas:</u> federated learning, distributed machine learning, imbalance data resampling, model bias reduction, advanced optimization algorithms
- <u>Publications:</u> 5+ research papers published or under review at venues like AAAI, ECML-PKDD, SSCI, and FC

BEng in Electrical Engineering, University of Electronic Science and Technology of China, China   2014-2018
- GPA: 3.28 (Top 10% with Scholarship for Outstanding Students throughout the program)

PROFESSIONAL EXPERIENCE

Research Associate, University of Birmingham, UK   Apr 2024 - Present
- Developed and analyzed machine learning models for tabular financial data, leveraging SVM, Random Forest, XGBoost, and Neural Networks
- Evaluated rebalance methods, including ROS, SMOTE, TabularGAN, and Triplet Oversampling, in a distributed environment
- Implemented a federated unlearning package based on PyTorch and Flwr frameworks
- Designed end-to-end pipelines incorporating data preprocessing, model training, parameter aggregation, unlearning, and evaluation
- Conducted performance benchmarking of unlearning methods across sample and feature dimensions
- Collaborated with cross-functional teams to integrate federated unlearning into existing AI systems

OPEN SOURCE PROJECTS

FedMind: Federated Learning Library for Researchers
- Federated learning library enabling fast prototyping and efficient evaluation of FL algorithms based on PyTorch
- Introduced custom data structures for module-level parallel tensor operations, simplifying algorithm development
- Supported multi-process and CUDA-enabled ML training with shared memory for efficient distributed simulations
- Included popular FL algorithms (e.g., FedAvg, FedProx, MFL, FedOpt) and facilitated easy addition of new algorithms
- Released 10+ versions on PyPI with over 7,000 downloads

HDF5-FEMNIST: FEMNIST Dataset in HDF5 Format
- Leveraged HDF5 format to store the FEMNIST dataset for efficient data loading and processing.
- Provided a PyTorch dataloader for seamless integration with federated learning frameworks.

jobSched: Slurm Parallel and Serial Job Scheduler in Python
- Extracted allocated resources from Slurm environment variables to determine host indices for distributed jobs.
- Automated parameter combination generation and job distribution across allocated resources.

PUBLICATIONS

Experimental Study of Class Imbalance in Federated Learning
(Oral Presentation) SSCI 2021

- Investigated the impact of class imbalance on federated learning model performance
- Proposed two metrics to quantify class imbalance from global and local perspectives

Triplets Oversampling for Class Imbalanced Federated Datasets
(Oral Presentation) ECML-PKDD 2023

- Designed a novel synthetic data generation algorithm for distributed class-imbalanced datasets.
- Generated informative synthetic samples by incorporating two majority neighbours for each minority sample
- Conducted extensive evaluations on federated learning benchmarks with multiple datasets and scenarios
- Proved robustness to small sample size, extreme class imbalance, and distributed private data sources

FedGA: Federated Learning with Gradient Alignment for Error Asymmetry Mitigation
Under Review, AAAI 2025

- Addressed error asymmetry in federated learning caused by class imbalance
- Connected error asymmetry to gradient misalignment and proposed label calibration techniques
- Demonstrated model convergence and performance improvements on heterogeneous datasets
- Proved model convergence and performance gain on heterogeneous datasets in 3 level evaluations

Benchmarking Federated Machine Unlearning Methods for Tabular Data
Under Review, FC 2025

- Evaluated federated unlearning on tabular data across sample and feature dimensions
- Implemented and benchmarked retraining, fine-tuning, and unlearning methods on real-world datasets

Rethinking the Initialization of Momentum in Federated Learning with Heterogeneous Data
Preprint, ArXiv

- Identified limitations in momentum aggregation for federated learning with heterogeneous data
- Improved model convergence using reversed accumulation weights for momentum

SKILLS

- **Programming Languages:** Python (Proficient), Rust, C++, JavaScript (Familiar)
- **Deep Learning Frameworks:** PyTorch, TensorFlow, Keras, ONNX, Scikit-learn, NumPy, Pandas
- **AI and Machine Learning:** NLP, LLMs, Federated Learning, Model Optimization, Data Heterogeneity
- **Tools and Technologies:** CUDA, Multiprocessing, Docker, gRPC, Git, Slurm
- **DevOps/MLOps:** CI/CD Pipelines, Automated Testing, Deployment, Monitoring, Logging
- **Mathematics and Algorithms:** Linear Algebra, Probability, Optimization, Algorithm Design