Yaopei Zeng

(86)13693476509 | 220019172@link.cuhk.edu.cn

EDUCATION

Huazhong University of Science and Technology
Qiming Honors College
Bachelor of Electronic Information Engineering

Wuhan, China
Sep 2016 - Jul 2020
GPA: 3.55/ 4.0

The Chinese University of Hong Kong, Shenzhen
School of Science an Engineering
MPhil of Computer and Information Engineering, Advisor: Baoyuan Wu

Shenzhen, China
Jan 2021 - Jun 2023
GPA: 3.50/ 4.0

Research Interests: Federated Learning, Long-tailed Learning

RESEARCH EXPERIENCE

Alibaba Group, Beijing | Research Intern   Apr 2022 – Jun 2023
- Exploring how to utilize abundant non-overlapping samples to enhance the vertical federated learning (VFL) model.
- Proposed an algorithm for transforming two parties' non-overlapping samples with the same labels into overlapping samples, which comprises pseudo-label prediction for non-overlapping samples in the passive party and canonical correlation analysis to measure representation similarities between non-overlapping samples from different parties.
- Implemented the vertical federated CTR prediction model with Pytorch, reproduced the existing VFL algorithm SFHTL, and compared to my algorithm over different proportions of overlapping samples.

Shenzhen Research Institute of Big Data, Shenzhen

Research Assistant | Federated Long-tailed Learning Project, Advisor: Prof. Baoyuan Wu   Oct 2021 – Jun 2023
- Researched the long-tailed class imbalance problem in federated learning (FL). One paper was accepted by ICCV 2023.
- Implemented various existing long-tailed learning re-balance strategies (CRT, LDAM, LADE, RIDE,...) from the centralized setting to FL, investigating the performance gap of re-balance in the centralized and FL.
- Proposed to use the magnitude of classifier' gradients uploaded by clients as the class prior to guide a multi-expert model training in FL to narrow the performance gap without data privacy leakage.
- The work is available at OpenReview: Label-distribution-agnostic Ensemble Learning on Federated Long-tailed Data.

Research Assistant | CCF-Hikvision Project, Advisor: Prof. Baoyuan Wu   Apr 2021 – Oct 2021
- Developed a FL platform, including both single node simulation and multi-node distributed training, which contained various CV, NLP, and medical datasets, and involved famous FL algorithms such as FedProx and SCAFFOLD.
- Added privacy guarantees to the platform, including differential privacy and homomorphic encryption algorithms.

Research Assistant | Advisor: Prof. Baoyuan Wu   Jan 2021 – Apr 2021
- Researched the influence of data heterogeneity (Non-IID) in FL, and found out that group normalization is more robust than batch normalization to the error when averaged the normalization layers' parameters of different clients.
- Designed a backdoor defense method of FL based on self-supervised learning.
- Involved in the AISecMatrix Program, the first overall framework focusing on the security issues of AI systems, working with Tencent Zhuque Lab.

Research Assistant | Advisor: Dr. Li Liu   Sep 2020 – Jan 2021
- Designed an privacy-preserving FL system for Cued Speech (a communication mode used with and among hard-of-hearing people) recognition, which received Fund of the National Natural Science Foundation of China.
- Proposed to use data mixup (a data augmentation method) to generate virtue samples for building a shared dataset on a server to address data heterogeneity in the FL system, and the virtue samples were encrypted with SMPC (a secure aggregation algorithm) before being uploaded to the server to protect privacy.

PROFESSIONAL EXPERIENCES

Intel, Shanghai | University-Industry Collaboration Project (remote)   Mar 2019 – Jun 2019
- Participated in building the 5G protocol stack on the DPDK (Data Plane Development Kit) platform.
- Implemented the packages receiving process simulation of BBU (Base band Unite) with C, which is a part 5G communication processes.

Sangfor, Changsha | Intern (remote)   Jun 2018 – Sep 2018
- Participated in developing a secure mailbox project provided for enterprises internal communication.
- Implemented the accounts module with C++, including creating and deleting an account in the mailbox and operating information associated with the account.

SKILLS AND PROFICIENCIES
- Programming : Python, C, C++, Matlab
- Interests: Photography, Ping-pong, Swimming.