香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr. TANG Zhenheng
Date	3 June 2024 (Monday)
Time:	10:00 am – 12:00 noon (35 mins presentation and 15 mins Q & A)
Venue:	1) DLB 637, 6/F, David C Lam Building, Shaw Campus 2) ZOOM (Meeting ID: 965 5550 6928) (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/bucs-reg (Deadline: 12:00 nn, 2 June 2024)

Collaborative Machine Learning at Scale: Efficiency and Generalization

Abstract

Recent years have seen profound progress in Machine Learning (ML) and Deep Learning (DL), such as GPT-3 and GPT4. Efficient and effective training of such large models requires the high quality and large quantity of training data, and dedicated distributed training methods with the enough computing resource. Federated Learning (FL) emerges as a solution to enhance training data quality and quantity, enabling users to collaboratively train DL models without compromising data privacy. In this thesis, we study how to improve the efficiency and generalization ability of FL. First, we understand how data heterogeneity influences FL performance at the intermediate feature level, finding out the feature shift phenomenon. We propose to exploit the shared noise data to calibrate shifted features, improving the convergence speed and the generalization performance. Second, we identify that client models contribute better generalization ability to the server model when having more similar data distribution. Based on this insight, we introduce to decouple the DL model, and train the high-level model with both original features and the shared estimated features. Third, considering the heterogeneous bandwidth problem, we develop a decentralized framework, GossipFL, aimed at optimizing communication efficiency and guaranteeing training convergence. Extensive experiment results demonstrate that our algorithms outperforms other SOTA FL methods over common used FL datasets. Finally, we envision collaborative machine learning to enrich computing resources by aggregating geo-distributed compute devices. We design a decentralized training system, FusionAI, which trains DL models using geo-distributed GPUs from different nodes. This work alleviates hardware scarcity in training large deep neural networks. A theoretical analysis shows the potential of FusionAI in efficient training large models with geo-distributed GPUs.

***** ALL INTERESTED ARE WELCOME *****