

TANG ZICHEN

+86 15854550175 ztangap@connect.hkust-gz.edu.cn github.com/trl730109

Education

The Hong Kong University of Science and Technology (Guangzhou)

Sep 2023 – Present

Master of Philosophy in Data Science and Analytics, GPA: 4.075/4.3

Guangzhou, China

Korea Advanced Institute of Science and Technology

Feb 2022 – Jun 2023

Exchange student in School of Computing

Daejeon, Korea

The Hong Kong University of Science and Technology

Sep 2019 – May 2023

Bachelor of Engineering in Computer Science and Mathematics (Double Major)

Hong Kong, China

Award

HKSAR Government Scholarship Fund - Reaching Out Award

June 2022

Publications

Conference

- **Z. Tang**, J. Huang, R. Yan, Y. Wang, Z. Tang, S. Shi, A. Zhou, X. Chu. Bandwidth-Aware and Overlap-Weighted Compression for Communication-Efficient Federated Learning. In the 53rd International Conference on Parallel Processing (ICPP'24).
- Z. Tang*, **Z. Tang***, J. Huang, R. Yan, Y. Wang, A. Zhou, S. Shi, B. Li, X. Chu. DreamDDP: Accelerating Distributed Training with Layer-wise Partial Synchronization. In the IEEE International Conference on Computer Communications (INFOCOM'25). (Under Review) *These authors contributed equally to this work.
- J. Huang, **Z. Tang**, R. Yan, Y. Feng, Z. Li, Z. Tang, A. Zhou, Y. Liang, X. Chu. Stale Information Matters: Efficiently Tackling Data Heterogeneity in Asynchronous Federated Learning with Model Calibration. In the IEEE International Conference on Computer Communications (INFOCOM'25). (Under Review)
- Z. Tang, J. Huang, **Z. Tang**, X. Kang, Y. Wang, P. Dong, S. Shi, X. Chu, B. Li. Capturing and Mitigating Gradient Aggregation Errors for Fault-Tolerant Distributed Training. In the Thirteenth International Conference on Learning Representations (ICLR'25) (Under Review)

Research Interest

- Federated Learning
- Distributed ML Systems
- LLM

Research Experience

Federated Learning

Sep 2023 – Jan 2024

Bandwidth-Aware and Overlap-Weighted Compression for Communication-Efficient Federated Learning *Guangzhou, China*

- Proposed a new scheduling method to dynamically adjust compression ratios based on heterogeneous client bandwidth, aligning model update transmission times to tackle communication bottlenecks and bandwidth heterogeneity.
- Discovered the heterogeneous overlap pattern of clients' compressed parameters and proposed a novel averaging method that utilizes a parameter mask to adjust the model updates in the aggregation process based on their occurrence frequency across clients.

Federated Learning

Dec 2023 – June 2024

Efficiently Tackling Data Heterogeneity in Asynchronous Federated Learning with Model Calibration

- Developed a novel semi-asynchronous FL method, K-FedQue, that leverages a dynamic model queue on the server side to utilize stale information from historical updates for model calibration, mitigating client drift caused by asynchronous updates and data heterogeneity.
- Formulated and analytically demonstrated K-FedQue achieves standard SGD convergence in non-convex settings.
- Proposed and implemented a staleness-aware regularization term on the local sub-problem to further counteract the effects of client drift. Extensive experiments showed that K-FedQue improves training efficiency by up to 1.76 times compared to SOTA asynchronous FL algorithms.

Distributed ML Systems

Feb 2024 – Aug 2024

DreamDDP: Accelerating Distributed Training with Layer-wise Partial Synchronization

Guangzhou, China

- Proposed partial synchronization that relaxes the strong synchronization in local SGD by layer-wisely decoupling model parameters into each iteration.
- Further introduced DreamDDP, which schedules synchronization according to the real-time profiled communication and computation time, maximizing communicated information with the minimal communication time by overlap.
- Theoretically proved that DreamDDP shows the same convergence rate as S-SGD by providing the convergence bounds. Experimental results tested on two GPU clusters with 32 GPUs demonstrated $1.49 - 3.91\times$ improvement over the SOTA algorithms, including local SGD and ASC-WFBP.

Projects

- Develop chatbots in Traditional Chinese Medicine domain | *Python*Jan 2024 – Present
- Finetuned LLM with Non-IID data sources to improve TCM-related query handling via Federated Learning.
 - Used RAG techniques to enhance the fine-tuned chatbot’s response accuracy and relevance.

Skills

Professional Skills: Python, Pytorch, PyTorch Distributed, Horovod, Transformers
Language Profficiency: English(TOEFL-105) Chinese(Native), Cantonese(Elementary)
Examination Performance: GRE 324, Verbal 154, Quantitive 170