Ziheng Qin

e0823059@u.nus.edu · (+86) 185-162-87789 · Ziheng

Education

National University of Singapore (NUS), Singapore 2021 – Present

PhD student in Computer Science, expected 2025

University of Southern California (USC), CA, USA 2020 – 2021

Master in Computer Science

University of Michigan, Ann Arbor (UM), MI, USA 2017 – 2019

B.S.E in Computer Science

Shanghai Jiao Tong University (SJTU), Shanghai, China 2015 – 2019

B.S.E in Electronic and Computer Engineering (ECE)

Experience

ICLR 2024 Oral (1.2%) 2022-2023

First Author Supervisor: Yang You

InfoBatch: Lossless Training Speed Up by Unbiased Dynamic Data Pruning

- Propose a novel framework to achieve lossless training acceleration by unbiased dynamic data pruning.
- Consistently obtains lossless training results on classification, semantic segmentation, vision pertaining (MAE and Diffusion Model), and instruction fine-tuning (LLaMA) tasks with 20% $\sim$ 40% saving.
- Compatible with various optimizers, coreset selection methods, and LoRA.

AAAI 2023 Outstanding Paper 2021-2022

Project Member Supervisor: Yang You

CowClip: reducing CTR prediction model training time from 12 hours to 10 minutes on 1 GPU

- This work identifies the bottleneck of large batch training in the recommendation system and solves it with the proposed CowClip algorithm, which greatly reduces the training time and cost.
- My contribution: help to analyze the large batch training bottleneck in the recommendation system, and participate in the algorithm discussion.

Google LLC. California, USA 2021 Summer

Summer Intern Manager: Yixin Shi

DLVM Cluster for VM-Based Distributed Training: A VM-based distributed training solution.

- Can create, setup, scale, execute commands, and manage files for a DLVM cluster with simple yaml file and command line.
- providing a choice with higher freedom and availability than the existing solution (TPU, GKE clusters etc).
- Can support different Machine learning frameworks and distributed learning frameworks.

Shanghai Jiao Tong Univesity & Intel Shanghai, China 2019 Summer

Project Member Instructor: Yong Long(SJTU), Forrest Zhao(Intel), Ruoyu Ying(Intel)

Cloud Gaming Architecture based on StarlingX and Akraino: An edge cloud gaming architecture.

- Using StarlingX and Kubernetes to build an edge server which do remote graphical computing in container for client during gaming with low latency, high compatibility and fast deployment.
- Attend Open Infrastructure Summit (Shanghai 2019) and do a presentation for this project as a speaker.
- Win Golden prize of SJTU-UMJI 2019 Summer Design Expo.

University of Michigan Student Research MI, USA 2018-2019

Project Leader Instructor: Prof. Michael Cafarella(UM)

Table-embeddings: trying to recover table column name from content.

- Using Pytorch, create hard-coded features together with Stanford-NER and neural network resulting in 90% accuracy prediction of 3000+ classes for web tables crawled from webpages.
- An early attempt to structurize unlabeled table data.

Skills

- Programming Languages: Python, C/C++, Java
- Platform: Linux
- Development: Machine Learning/Reinforcement Learning/Deep Learning/Data Mining/Distributed Training, MySQL, Spark, XGBoost, CSS, JavaScript, Html, Hadoop, MATLAB, Mathematica, PPT, LaTeX

Miscellaneous

- GitHub: https://github.com/henryqin1997
- Google Scholar: https://scholar.google.com/citations?user=I04VhPMAAAAJ
- Languages: English, Chinese(Mandarin)