Wentao Guo

Homepage | GitHub | LinkedIn | wg247@cornell.edu

RESEARCH INTEREST

Model and Data Efficiency of Foundation Models, Distributed Learning, Machine Learning Systems

EDUCATION

Cornell University Jun 2022 - Dec 2023

Master of Engineering in Computer Science, GPA: 4.031

Cornell University Sep 2018 - May 2022

B.S. in Computer Science with Honors, Magna Cum Laude, GPA: 3.890

PUBLICATION & MANUSCRIPT

(* denotes equal contribution.)

- A. Feder Cooper*, **Wentao Guo***, Khiem Pham*, Tiancheng Yuan, Charlie F. Ruan, Yucheng Lu, Christopher De Sa. **"CD-GraB: Coordinating Distributed Example Orders for Provably Accelerated Training."** *In proceedings of NeurIPS'23.* [paper] [poster]
- Yucheng Lu, **Wentao Guo**, and Christopher De Sa. **"GraB: Finding Provably Better Data Permutations than Random Reshuffling."** *In proceedings of NeurIPS'22.* [paper] [poster]
- **Wentao Guo***, Andrew Wang*, Bradon Thymes, Thorsten Joachicms. **"Ranking with Slot Constraints."** [paper]
- Tao Yu*, **Wentao Guo***, Jianan Canal Li*, Tiancheng Yuan*, Christopher De Sa. **"MCTensor: A High-Precision Deep Learning Library with Multi-Component Floating-Point."** In *Hardware Aware Efficient Training (HAET) workshop at ICML'22.* [paper] [poster] [code] [video]

RESEARCH EXPERIENCE

Cornell University Jun 2021 - Present

Research Assistant, Prof. Christopher De Sa's Lab, Cornell University

- **CD-GraB Project: find a good distributed data ordering with decentralized data**
  - Developed CD-GraB algorithm that enjoys a linear speedup of convergence rate on the number of workers and achieves provably better convergence rate than distributed random reshuffling (D-RR).
  - Demonstrated both iteration-wise and wall-clock time convergence speedup over D-RR.
  - The paper is accepted by **NeurIPS'23** main track and also DMLR workshop in ICML'23.
- **GraB Project: find a good data ordering for SGD with centralized data**
  - Collaborated to develop GraB algorithm that balances the gradients of each example to find a better data ordering than random reshuffling (RR).
  - Demonstrated both iteration-wise and wall-clock time convergence speedup over RR.
  - The paper was presented in **NeurIPS'22**.
- **MCTensor Project: efficient high-precision arithmetic with multi-component floats**
  - Developed the MCTensor library that enables efficient high-precision floating-point arithmetic with multi-component low-precision floats, and implemented basic arithmetic algorithms and operators, and the high-level NN modules and optimizers that mirrored PyTorch library structures.
  - Demonstrated that the performance of MCTensor models in 16-bit can match the 32-bit weights in hyperbolic learning tasks.
  - The paper was presented in **HAET workshop** at ICML'22.

Cornell University Jun 2022 - May 2023

Research Assistant, Prof. Thorsten Joachicms's Lab, Cornell University

- **MatchRank Project: ranking with slot constraints**
  - Investigated the ranking problem under slot constraints and formulated the ranking objective as the size of maximum bipartite matching (MBM) on sampled candidate-slot bipartite graphs.
  - Developed the MatchRank algorithm as a greedy algorithm on submodular monotone objective, and further optimized the time complexity of the MatchRank algorithm on both greedy query and MBM finding augmenting paths side.
  - Generalized the MatchRank algorithm on binary relevance bipartite graph (admission problem) to continuous-valued relevance bipartite graph (general recommendation problem).
  - Performed experiments on Mulan binary multilabel datasets, Cornell undergraduate admission dataset, and Amazon recommendation datasets.

ENGINEERING EXPERIENCE

- **Developer Lead**
  Pathways Project, Prof. René Kizilcec's Lab, Cornell University Jun 2021 - May 2023
  - Developed the backend with Flask, MongoDB, and Redis, designed search algorithms that provided diverse suggestions on course enrollment choices, and iterated algorithms from students' feedback.
  - Deployed and maintained the website to serve more than 3000 Cornell students.

- **Backend Developer & Tester Lead**
  Course Management System, Cornell University Sep 2019 - May 2022
  - Fixed 10s MySQL and Java production bugs on backend, created 75 and reviewed 76 peer's pull requests, and supervised new members and held weekly meetings to manage the team.
  - The website serves more than 8000 students in over 100 courses in Cornell University.

- **Game Development Intern**
  QQ Speed Mobile Team, Tencent, Shenzhen, China Jun 2020 - Aug 2020
  - Programmed game modules in Unity with C#, created tools to accelerate project loading and compilation time, and analyzed the performance of C# libraries on serialization and deserialization.

TEACHING EXPERIENCE

- **CS 4787 Principles of Large-Scale Machine Learning Systems** Fall 2023
- **CS 4780 Intro to Machine Learning** Spring 2023
- **CS 3110 Data Structures & Functional Programming** Fall 2021

ACADEMIC SERVICE

- **NeurIPS'23, ICLR'24 Reviewer**

HONOR

Cornell Engineering Honor Society (Tau Beta Pi), Dean's List for 6 semesters, Honorable Mention in MCM 2018