Rui Pan

https://ruipan.xyz/ | rpan33@wisc.edu | (+1) 608-960-0303

EDUCATION

University of Wisconsin-Madison   Madison, WI, USA
Bachelor of Science double majoring in Computer Science and Mathematics   *Sep 2018 – Dec 2021 (Expected)*
- GPA: 3.96/4.00

RESEARCH INTERESTS

I am broadly interested in big data systems (Machine Learning Systems, Cloud/Distributed Systems, Networks). For now, I am doing research on **Machine Learning Systems**, with a focus on optimizing systems for Deep Neural Network training. I am applying to Ph.D. programs that start in Fall 2022!

PUBLICATIONS

[**1**] [Conference Paper] Fair and Efficient Cluster Scheduling for Dynamic Adaptation in Machine Learning. Pengfei Zheng, **Rui Pan**, Tarannum Khan, Shivaram Venkataraman, Aditya Akella. 2021. *In submission.*

[**2**] [Poster] AgDH: A System for Gathering and Distributing Dairy Data. **Rui Pan**, Steven R. Wangen, Michael C. Ferris. 2020. *Presented at the 3rd Annual WID Symposium.*

RESEARCH EXPERIENCE

Undergraduate Research Assistant w/ Prof. Shivaram Venkataraman   Madison, WI, USA
Project 1: Fair and Efficient Resource Allocation for DNN Training in GPU Clusters   *Mar 2021 – Present*
- In this work, we develop a policy to co-optimize long-term fairness and efficiency of the scheduling/resource allocation of resource-adaptive deep learning training workloads in large-scale multi-tenant GPU clusters.
- I contributed to the implementation and integration of our novel allocation policy into Gavel [OSDI '20], an existing scheduling framework. I also implemented in Gavel the mechanism to support dynamic adaptation (e.g., batch size scaling) of training workloads.
- I contributed to the implementation of dynamic optimizations, Accordion [MLSys '21] & Gradient Noise Scale [arXiv '18], for common DNN training workloads to increase the system throughput without loss of accuracy.
- On a trace of real-world workloads, we outperform state-of-the-art scheduling policies (Themis, Gavel, AlloX) in efficiency by 35% and fairness by 75% at the same time.

Project 2: How Structured Backpropagation Pruning Improves Deep Learning Clusters   *Jun 2020 – Feb 2021*
- In this work, we systematically control the amount of backpropagation at individual workers in distributed DNN training. This technique, Structured Backpropagation Pruning, simultaneously reduces network bandwidth, compute utilization, and memory use while preserving model quality.
- I developed an iteration-level cluster scheduler based on existing frameworks such as PyTorch Elastic and BytePS [OSDI '20] to support fine-grained scheduling, frequent checkpointing, different communication protocols (all_reduce/parameter server), and constant worker migration with a low overhead in order to capitalize on the resources saved by Structured Backpropagation Pruning.

Undergraduate Research Assistant   Jan 2020 – Mar 2021
Advisors: Dr. Steven Wangen and Prof. Michael Ferris   *Madison, WI, USA*
- In this work, we propose Dairy Brain, an analytics platform for evaluating and predicting the performance of dairy cows by aggregating large quantities of dairy data.
- I developed, deployed and maintained the Agricultural Data Hub (AgDH), a warehouse-scale computing infrastructure for the collection, storage, homogenization, and distribution of dairy farm's feeding, milking, and management data in a series of PostgreSQL data marts.
- I hosted poster sessions at the 3rd WID Research Symposium and presented in outreach meetings for the local dairy industry.

Incoming Research Intern @ Max Planck Institute for Informatics   Feb 2022 – May 2022
Advisor: Prof. Yiting Xia (Network-accelerated distributed machine learning)   *Saarbrücken, Germany*

Professional Experience

Tutor in Computer Science and Math
Sep 2019 – May 2020

Undergraduate Learning Center @ College of Engineering, UW-Madison
Madison, WI, USA

- Held weekly drop-in and by-appointment tutoring sessions to help students with understanding concepts and debugging programming assignments.
- Tutored students in 8 core introductory computer sciences and 5 math courses, and offered information and personal advice on course enrollment and career choices.
- Helped a total of ∼100 students.

Subtitle Translator
May 2019 – May 2020

Coursera
Online

- Volunteered in translating English subtitles to Simplified Chinese subtitles in multiple courses (An Introduction to Programming, Building Web Applications in PHP, Neural Networks and Deep Learning).

Student Instructor
Sep 2017 – Aug 2018

Research and Development Center @ High School Afiiliated to Shanghai Jiao Tong University
Shanghai, China

- Designed two introductory computer science courses (Intro to AP CS & Data Analysis with Python).
- Adapted course material from CS61A & CS61B @ Berkeley and CS50 @ Harvard.
- Gave lectures to 10+ students.

Relevant Courses

Graduate-level: Advanced Operating Systems, Big Data Systems (audit I/P), High Performance Computing
Undergraduate CS: Algorithms, Artificial Intelligence, Bioinformatics, Computer Architecture, Computer Vision, Database Systems, Data Analysis, Networks, Operating Systems, Theory of Computing
Undergraduate Math: Calculus, Combinatorics, Discrete Math, Mathematical Data Science (I/P), Numerical Linear Algebra, Probability
Coursera: Cloud Computing (I/P), Deep Learning, Software Engineering, Front-end Web Design, Back-end Web Applications, Cryptography

Technical Skills

Languages: Python, Java/C#, C/C++, SQL, JavaScript, HTML/CSS, R
Frameworks and Tools: PyTorch, Git, Docker, PostgreSQL, CUDA, OpenMP, MPI, Apache Spark

Honors and Awards

- 2022 Research Fellowship, Max Planck Institute for Informatics (MPI-INF), Internet Architecture Group

- 2021 Cornell, Maryland, Max Planck Pre-doctoral Research School (CMMRS)

- 2018 Dean's List (6 semesters)