JIANRU DING

CHICAGO, 60615, UNITED STATES 6149068052

DETAILS

Chicago, 60615, United States
jrding@uchicago.edu

TECHNICAL SKILLS

Skills: Kernel & Low-level development, Parallel computing, Data mining, Machine learning

Languages: Python, C, C++, Java, SQL

Frameworks & APIs: Apache Spark, Tensorflow

TEACHING EXPERIENCE

Teaching Assistant
University of Chicago
September 2020 — May 2021
Parallel Computing
Computer Architecture for Scientists
Computer Architecture

Teaching Assistant
The Ohio State University
January 2018 — September 2019
Intro to Database System
Principle of Programming Languages
Systems I: Introduction to Low-Level Programming and Computer Organization

INDEPENDENT PROJECTS

Occupant Wellbeing Project
Developed an Electrocardiography analyzing system using deep learning to detect driver emotions in a Honda R&D sponsored project

Decoupled Neural Interfaces for Residual Neural Network using Synthetic Gradients
Developed an RNN-based decoupled neural interface that takes residues into its prorogation prediction model reducing GPU memory usage by roughly 50%

X86 Processor
Implemented a 32-bit 5-staged x86 processer in C as part of a course project

EDUCATION

Ph.D. in Computer Science — **University of Chicago, Chicago**
September 2020 — Current
Advisor: Dr. Henry Hoffman
Research Interest: My research interest lies in the joint of HPC, Computer Architecture and Operating Systems, and Machine Learning. My work focuses on control systems that adapt computing resource management to large-scale workload fluctuations to meet high-level user-defined goals.

B.S. in Computer Science — **The Ohio State University, Columbus**
B.S.B.A. in Finance
August 2016 — August 2020
Honor Engineering Program
Honor Thesis: Characterizing Service Level Objectives
Graduated with Cum Laude Honor

INDUSTRY EXPERIENCE

Intern Pre-Sales Tech Consultant at USTC iFLYTEK Science and Technology Co., Ltd., Hefei, China
May 2019 — August 2019

- Developed the customized deep learning-based speech detection system for various client companies as part of a team
- The speech detection system reaches keyword and semantic detection accuracy of more than 99% for clients with expertise in different fields
- The final system was actively adopted by several companies

RESEARCH PROJECTS

University of Chicago, Advisor: Prof. Henry Hoffmann

The UChicago Updown Project
June 2022 — Current

- Proposed the first-of-its-kind million-scale distributed task load-balancer (ongoing)
- Designed and implemented fine-grained distributed training & inference for GCN
- Leader of a team of 11 members on assembler implementation and testing
- Co-designer of the load balancing scheme for KV Map Shuffle Reduce (ICPC ’23)

DPS: Adaptive Power Management for Overprovisioned Systems (SC ’23)
September 2020 — April 2023

- Designed the first-of-its-kind model-free stateful power management system for overprovisioned clusters, which yields close performance to optimal model-based approaches and outperforms SLURM by up to 12.4%
- Developed and released the power management program as open source

The Ohio State University, Advisor: Dr. Christopher Stewart

Cache-based Computational Sprinting (SoCC '18)
April 2018 — June 2020

- Developed a CNN and gcForest based Service Level Objective (SLO) computational sprinting modeling approach increasing concurrent cache usage and throughput
- The approach reduced slack between SLOs and application latency from 20% to 1%

Characterizing Service Level Objectives (ICAC '19)
August 2018 — January 2019

- Designed a well-defined repeatable Systematic Literature Review (SLR) process for data mining Service Level Objectives (SLO) that reduces potential bias within large-amount literature reviews
- Applied the SLR to accumulate more than 80 sets of Service Level Objective (SLO) samples by datamining more than 50 industrial products and 9,500 published articles

PUBLICATION LIST:

Rajasukumar, A., Su, J., Wang, Y., Su, T., Nourian, M., Diaz, J.M., Zhang, T., Ding, J., Wang, W., Zhang, Z., Jeje, M., Hoffmann, H., Li, Y., & Chien, A.A. (2024). UpDown: Programmable fine-grained Events for Scalable Performance on Irregular Applications. *ArXiv, abs/2407.20773*.

Wang, Y., Rajasukumar, A., Su, T., Nourian, M., Diaz, J. M. M., Pervaiz, A., Ding, J., Colley, C., Wang, W., Li, Y., Gleich, D., Hoffmann, H.,& Chien, A. A. Efficiently Exploiting Irregular Parallelism Using Keys at Scale.

Ding, J., & Hoffmann, H. (2023, November). DPS: Adaptive Power Management for Overprovisioned Systems. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 1-14).

Ding, J. (2020). *Characterizing Service Level Objectives for Cloud Services: Motivation of Short-Term Cache Allocation Performance Modeling* (Honor B.S. Dissertation. The Ohio State University).

Ding, J., Cao, R., Saravanan, I., Morris, N., & Stewart, C. (2019, June). Characterizing service level objectives for cloud services: Realities and myths. In *2019 IEEE International Conference on Autonomic Computing (ICAC)* (pp. 200-206). IEEE.

Nathaniel Morris, Indrajeet Saravanan, Pollyanna Cao, Jerry Ding, and Christopher Stewart. 2018. SLO Computational Sprinting. In Proceedings of the ACM Symposium on Cloud Computing (SoCC '18). Association for Computing Machinery, New York, NY, USA, 510. https://doi.org/10.1145/3267809.3275452