

Yunjia Zhang

Email: yunjia@cs.wisc.edu

Mobile: (608) 338-5063

Homepage: pages.cs.wisc.edu/~yunjia/

EDUCATION

University of Wisconsin-Madison, Department of Computer Science

Madison, WI

Program: *Ph.D.* in Computer Science

Sep. 2019 – May. 2024 (expected)

- Advisors: Prof. Jignesh M. Patel (currently at CMU), Theodoros Rekatsinas (currently at Apple)
- Collaboration: Research assistant at Microsoft Gray Systems Lab (GSL), mentor: Avrielia Floratou
- Thesis Title: Machine Learning for Data Analytics: Profiling, Querying, and Beyond
- Services: Teaching assistant for CS 400 Programming III and CS 564 Database Management Systems

Wuhan University, Department of Computer Science

Wuhan, China

Program: *B.Eng.* in Computer Science and Technology

Sep. 2015 – Jun. 2019

- Overall GPA: 3.92/4.00 (Major GPA: 3.97/4.00, Overall rank: 1/164)
- Advisor: Prof. Hao Huang
- Research Focus: Diffusion network recovery

PUBLICATIONS

- **GPTuner: A Manual-Reading Database Tuning System via GPT-Guided Bayesian Optimization**
Jiale Lao, Yibo Wang, Yufei Li, Jianping Wang, **Yunjia Zhang**, Zhiyuan Cheng, Wanghu Chen, Mingjie Tang, Jianguo Wang
arXiv preprint, under submission.
- **ReAcTable: Enhancing ReAct for Table Question Answering**
Yunjia Zhang, Jordan Henkel, Avrielia Floratou, Joyce Cahoon, Shaleen Deep, Jignesh M. Patel
arXiv preprint, under submission. (In collaboration with Microsoft GSL)
- **Simple Adaptive Query Processing vs. Learned Query Optimizers: Observations and Analysis**
Yunjia Zhang, Yannis Chronis, Jignesh M. Patel, Theodoros Rekatsinas
International Conference on Very Large Databases (VLDB '23)
- **Schema Matching using Pre-Trained Language Models**
Yunjia Zhang, Avrielia Floratou, Joyce Cahoon, Subru Krishnan, Andreas Mueller, Dalitso Banda, Fotis Psallidas, Jignesh M. Patel
International Conference on Data Engineering (ICDE '23) (In collaboration with Microsoft GSL)
- **Can Transfer Learning be used to build Query Optimizers?**
Yunjia Zhang, Yannis Chronis, Jignesh M. Patel, Theodoros Rekatsinas
The Conference on Innovative Data System Research (CIDR '22)
- **A Statistical Perspective on Discovering Functional Dependencies in Noisy Data**
Yunjia Zhang, Zhihan Guo, Theodoros Rekatsinas
ACM SIGMOD International Conference on Management of Data (SIGMOD '20)
- **Statistical Estimation of Diffusion Network Topologies**
Keqi Han, Yuan Tian, **Yunjia Zhang**, Ling Han, Hao Huang, Yunjun Gao
International Conference on Data Engineering (ICDE '20)
- **Fast Inference of Diffusion Networks without Infection Temporal Information**
Yueming Sun, **Yunjia Zhang**, Qian Yan, Lu Chen, Hao Huang, Yunjun Gao
National Database Conference, National Data Buoy Center (NDBC '18)

PROFESSIONAL EXPERIENCE

- May. 2021 – Present* Research assistant at Microsoft Gray Systems Lab (GSL). Mentor: Avrielia Floratou.
- May. 2020 – Present* Research assistant at Database Research Group, University of Wisconsin-Madison.
- Sep. 2019 – May. 2020* Teaching assistant at Department of Computer Science, University of Wisconsin-Madison.

Nov. 2018 – Feb. 2019 Research intern at Database System Research Lab, NUS, Singapore.

Jul. 2018 – Oct. 2018 Research intern at Singtel Cognitive and AI Lab, NTU, Singapore.

Sep. 2017 – Jun. 2018 Research assistant at State's Key Lab of Software Engineering, Wuhan University, China.

RESEARCH HIGHLIGHTS

Gray Systems Lab, Microsoft

Madison, WI

Research Assistant

Advisor: Avriella Floratou

May. 2021 – May. 2022

- **Research on Relational Schema Mapping using Large Language Model**

- Designed a modular featurizer to encode the attribute pairs, which includes a fine-tuned language model-based sub-featurizer. The LM-based featurizer is based on BERT and fine-tuned on the industry schema in Azure.
- Applied semi-supervised framework with a neural network meta-learner in the schema mapping model.
- Experiments showed that our model reduces the matching labels for real customer schema by up to 81%.
- Finished paper: *Schema Mapping via Language Models (ICDE '23)*. Patent application in process.

Database Research Group, University of Wisconsin-Madison

Madison, WI

Research Assistant

Advisors: Prof. Jignesh M. Patel, Theodoros Rekatsinas

Mar. 2022 – May. 2023

- **Research on Adaptive Query Processing vs. Learned Query Optimizers**

- Analyzed learned query optimizers and verified that: 1) join order contributes more to the performance of learned QOs, and 2) learned QOs may not generalize well to completely unseen queries.
- Proposed to use LIP+AJA as an adaptive query processing method and implemented them into PostgreSQL
- Experimentally evaluated LIP+AJA against modern learned query optimizers and showed LIP+AJA is comparable to modern learned QOs and even better in many cases.
- Finished paper: Simple Adaptive Query processing vs. Learned Query Optimizers: Observations and *analysis*, (*VLDB '23*).

- **Research on Analyzing and Transferring Query Optimizers**

Mar. 2020 – May. 2021

- Proposed the problem of transferring query optimizers: learning from one optimizer and apply the learnt model on another data system without an advanced query optimizer or without optimizers.
- Performed preliminary experiments on transferring query optimizer, and it showed an execution time improvement by up to 50x, transferring from PostgreSQL to Couchbase.
- Finished paper: *Can Transfer Learning be used to build Query Optimizers? (CIDR '22)*.

- **Research on Functional Dependency Discovery**

Sept. 2019 - Mar. 2020

- Applied structure learning method on dependency discovery (FDX), which reduces the complexity to quadratic to the number of attributes.
- Experiments on FDX showed that FDX outperforms the best state-of-the-art baseline by up to 88% in F1-score.
- Finished paper: *A Statistical Perspective on Discovering Functional Dependencies in Noisy Data (SIGMOD '20)*.

HONORS & AWARDS

- **Outstanding academic award** with full tuition fee remission (awarded for top 0.5%) of Wuhan University
- **Outstanding Undergraduate** (awarded for top 3%) of Wuhan University
- **Meritorious Winner** for MCM/ICM (awarded for top 9%)
- **Merit Student** (awarded for top 4%) of Wuhan University for three consecutive academic years

PROJECTS & COURSE DESIGNS

- **Multi-terminal gymnastic management system:** designed a system architecture based on the MVC design pattern and implemented server application using LARAVEL framework.
- **Embedded intelligent monitoring system:** constructed an embedded system on raspberry pi for intelligent home monitoring with facial recognition, remote warning, and gesture control.

PROGRAMMING SKILLS

- **Programming Languages:** Python, SQL, C/C++, Java