

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 Sciences

Madison, WI, US

Program: Ph.D. in Computer Science

Sept. 2019 – Present

- Research Focus: Database
- Advisor: Theodoros Rekatsinas, Jignesh Patel (Database Group)
- Teaching Assistant: Programming III (Data structures in Java), Database Management Systems

Wuhan University, School of Computer Science

Wuhan, China

Program: B.Eng. in Computer Science and Technology

Sept. 2015 – Jun. 2019

- Overall GPA: 3.92/4.0 (Major GPA: 3.97/4.0, Overall rank: 1/164)
- Advisor: Hao Huang
- Research Focus: data mining, diffusion networks

PUBLICATIONS

- ***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), 06/2020

- ***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), 04/2020

- ***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), 05/2018

RESEARCH EXPERIENCE

Database Research Group, University of Wisconsin-Madison

Madison, WI

Research Assistant

Advisor: Prof. Theodoros Rekatsinas, Jignesh Patel

- **Research on Interpretable Query Optimizer for Relational DBMS**

Mar. 2020 - Present

- Used functional dependency discovered by FDX to get a better cardinality estimation, and plugin cardinalities into Postgresql to perform end-to-end test
- Proved that a simple set of execution plan generation rules plus a plan cost model (Postgresql) is enough to get good plans

- **Research on Functional Dependency Discovery**

Oct. 2019 - Mar. 2020

- Applied structure learning method on dependency discovery (FDX), which reduce the complexity to quadratic to the number of attributes
- Performed extensive experiments on FDX with state-of-the-art baselines, which proved FDX outperforms those
- Finished paper *A Statistical Perspective on Discovering Functional Dependencies in Noisy Data*, which showed up in SIGMOD'20

Database System Research Lab, National University of Singapore

Singapore

Research Assistant

Advisor: Prof. Xiaokui XIAO

- **Research on Parallel Computation of LSH for Machine Learning**

Nov. 2018 - Feb. 2019

- Comprehensive literature review on locality sensitive hashing used in machine learning to reduce computation
- Designed a PyTorch module to implement the algorithm and tested its efficiency and accuracy with real-world datasets

Singtel Cognitive and AI Lab for Enterprises, Nanyang Technological University

Singapore

Research Assistant

Advisor: Prof. Gao CONG

Jul. 2018 – Oct. 2018

• Recommender Systems for Further Education

- Proposed EduLinkedIn - a two-stage word embedding model which could deal with complete cold start to recommend further education positions to candidates
- Designed a crawler to semi-automatically generate synthetic datasets for preliminary experiments, and implemented word2vec, doc2vec and other basic modules using TensorFlow

State's Key Lab of Software Engineering, Wuhan University

Wuhan, China

Research Assistant

Advisor: Prof. Hao HUANG

Jan. 2018 - Jun. 2018

• Research on Reconstruction of Bayesian Network without Temporal Information

- Established a novel EM-based model to reconstruct the Bayesian Network and theoretically proved its efficiency
- Conducted experiments which indicated that our algorithm improved more than 20% in accuracy compared with previous works including NetRate, InfoPath etc.
- Completed paper *Fast Inference of Diffusion Networks without Infection Temporal Information*

Institute of Software, Chinese Academy of Sciences

Beijing, China

Research Assistant

Advisor: Prof. Limin GUO

Jul. 2017 - Aug. 2017

• Research on Efficient Location-based Searching Algorithm

- Investigated algorithms used in range-query and k nearest neighbors (kNN) searching, and performed them to deal with million data volume
- Developed a grid-based algorithm, which can accelerate the range-query searching procedure
- Designed an algorithm aiming at searching k nearest neighbors (kNN) using improved heap theory and proved its efficiency under bulk requests of nearest query

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

- **Crowd-sourcing service platform:** developed an Android platform to provide crowd-sourcing service, using the Client-Server mode
- **Patent information analyzing platform:** built a platform for searching and analyzing patent information using Java and PHP
- **GIS prototype system using QGIS developing API:** constructed a GIS prototype system providing basic functions of map controlling with C++ and QT
- **3D model displaying and controlling software:** constructed a 3D model displaying and controlling software providing functions like 3D picking with C++ and OpenGL
- **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
- **Real time express information reminder:** built a real time express information reminder on mobile devices by extracting online information using java and python crawlers

PROGRAMMING SKILLS

- **Languages:** Python, Java, C/C++, SQL, MATLAB
- **Environments & IDEs:** Linux, Windows, PyCharm, Visual Studio, Eclipse