JINYANG LI

Computer Science and Engineering, University of Michigan, Michigan, USA
jinyli@umich.edu | https://JinyangLi01.github.io

EDUCATION

University of Michigan, Ann Arbor — Michigan, USA
Ph.D. candidate, Computer Science and Engineering — Sep. 2020 – Present
Advisor: H. V. Jagadish

University of Michigan, Ann Arbor — Michigan, USA
Master of Science, Computer Science and Engineering — Sep. 2020 – Jul. 2023
Advisor: H. V. Jagadish

Peking University — Beijing, China
Bachelor of Science, Computer Science and Technology — Sep. 2016 – Jun. 2020

RESEARCH INTEREST

- Query processing and optimization.
- Responsible data management: fairness, bias, and diversity issues.

PUBLICATIONS

1. **Jinyang Li**, Yuval Moskovitch, Julia Stoyanovich, H. V. Jagadish
   Query Refinement for Diversity Constraint Satisfaction
   VLDB 2024
   PDF
2. **Jinyang Li**, Alon Silberstein, Yuval Moskovitch, Julia Stoyanovich, H. V. Jagadish
   ERICA: Query Refinement for Diversity Constraint Satisfaction
   VLDB Demo 2023
   Demo Award honorable mention
   PDF
3. Yuval Moskovitch, **Jinyang Li**, H. V. Jagadish
   Dexer: Detecting and Explaining Biased Representation in Ranking
   SIGMOD Demo 2023
   PDF
4. **Jinyang Li**, Yuval Moskovitch, H. V. Jagadish
   Detection of Groups with Biased Representation in Ranking
   ICDE 2023
   PDF
5. Yuval Moskovitch, **Jinyang Li**, H. V. Jagadish
   Bias analysis and mitigation in data-driven tools using provenance
   Proceedings of the 14th International Workshop on the Theory and Practice of Provenance, 2022
   PDF
6. **Jinyang Li**, Yuval Moskovitch, H. V. Jagadish
   DENOUNCER: detection of unfairness in classifiers
   VLDB Demo 2021
   PDF
7. Yinda Zhang, **Jinyang Li**, Yutian Lei, Tong Yang, Zhetao Li, Gong Zhang, Bin Cui
   On-Off Sketch: A Fast and Accurate Sketch on Persistence
   VLDB 2021
   PDF
8. Tong Yang, Haowei Zhang, **Jinyang Li**, Junzhi Gong, Steve Uhlig, Shigang Chen, Xiaoming Li, HeavyKeeper: An Accurate Algorithm for Finding Top-k Elephant Flows
   IEEE/ACM Transactions on Networking (ToN), 2019
   PDF

RESEARCH EXPERIENCE

Research Assistant, Database Group, University of Michigan Michigan, USA
Advisor: Professor H. V. Jagadish Sep. 2020 – Present

- We study the problem of modifying relational queries to have the result satisfy constraints on the sizes of multiple subgroups in it to improve diversity and group representation in query results With the help of a provenance model, we develop an efficient query refinement algorithm. (Pub. 1, 2)
- We study the problem of detecting groups with biased representation in the top-k-ranked items, eliminating the need to pre-define protected groups. We propose efficient search algorithms for two different fairness measures: global bound representation, and proportional representation. We also propose a method to explain the bias in the representations of groups utilizing the notion of Shapley values. (Pub. 3, 4, 5, 6)
- We study the problem of coverage and bias in image datasets by examining the activation level of a CNN model. The lack of activation in particular neurons would indicate coverage gaps, which can be translated into meaningful insights through advanced feature visualization techniques. (Ongoing work)
- We study how to measure fairness metrics of time series in non-stationary environments by applying an exponential time decay to traditional fairness metrics. We design new algorithms to monitor fairness status and flag anomalies in real time. (Ongoing work)

Applied Scientist Intern, AIRE, Amazon.com Inc. California, USA
Manager: Huzefa Rangwala May 2024 – Aug. 2024

- We studied the problem of automatic SQL generation for table transformation, to generate SQL for the table transformation given well-defined input and output table schema.
- We use the chain-of-thoughts approach to guide the Large Langrage Model (LLM) and to decompose complex tasks into small subtasks that can be handled by LLM.
- We use LLM to generate the DAG of the table transformation task using pre-defined SQL operations and templates, and then instantiate each template.

Research Intern, Infrastructure Lab, ByteDance Inc. California, USA
Manager: Ron Hu Jun. 2023 – Aug. 2023

- We applied the fairness metrics and algorithms from my paper (Pub. 4) to TikTok's hashtag page rankings to detect potential representation issues and analyze the possible causes.
- To analyze possible bias introduced by TikTok's algorithm, we use fairness metric ranking-based equal opportunity in TikTok's for-you-page recommendation with users' interests considered and compare it with that from the random traffic.

Research Assistant, FORWARD Data Lab, UIUC Illinois, USA
Advisor: Professor Kevin Chen-Chuan Chang Jun. 2019 – Jul. 2020

- We studied ordered access for relational data. We propose a theoretical optimality that takes the number of updates into consideration and theoretically proves the lower bound of change-aware order indexing.
- We designed a novel index structure realizing the theoretical optimality in an update-aware manner to support ordered access to RDBMS.

Research Assistant, Network Big Data Lab, Peking University Beijing, China
Advisor: Professor Tong Yang Jan. 2019 – Jun. 2019

- We worked on algorithms and data structures to find top-k elephant flows in network traffic measurement. The proposed algorithm incurs a small, constant processing overhead per packet. (pub. 8)
- We study the problem of the persistence of items -- whether an item appears recurrently in many time windows of a data stream. We use the characteristic that the persistence of an item is increased periodically to compress increments to accurately estimate persistence and find persistent items. (Pub. 7)

TECHNICAL SKILLS

Programming languages: Python, C, C++, GO, SQL
Data platforms: MongoDB, Apache Hadoop, LLM