370 Jay Street
Brooklyn, NY 11201
hp1326@nyu.edu
Webpage : haoyueping.github.io/
Github : github.com/haoyueping

Haoyue Ping

PhD candidate, Computer Science, NYU Tandon

Research Interests

- Develop new techniques for preference modeling and inference
- Build probabilistic databases for preference data
- Boost data sharing with differential privacy
- Solve real world problems with data science

Education

New York University *Jan' 2019 - Jun' 2022 (expected)*
Ph.D. candidate in Computer Science, Tandon School of Engineering, GPA: 4.0
Advised by Dr. Julia Stoyanovich
Research Assistant in Visualization and Data Analytics Research Center (VIDA)
Thesis: *Querying Election Databases: Bridging Probabilistic Preferences and Social Choice*
Committee: Julia Stoyanovich (chair), Juliana Freire, Benny Kimelfeld, and Torsten Suel

Drexel University (transferred) *Sep' 2014 - Dec'2018*
Ph.D. student in Computer Science, GPA: 3.88
Advised by Dr. Julia Stoyanovich
Research Assistant in Drexel Database Group

University of Chinese Academy of Sciences *Sep' 2011 - Jul' 2014*
M.Eng. in Electronics and Communications Engineering, GPA: 3.14
Advised by Dr. Zhan Zhao
Research Assistant in State Key Laboratory of Transducer Technology

University of Electronic Science and Technology of China *Sep' 2007 - Jul' 2011*
B.S. in Biomedical Engineering, GPA: 3.38

Publications

1. **Haoyue Ping**, Julia Stoyanovich, and Benny Kimelfeld. 2020. Supporting hard queries over probabilistic preferences. Proc. VLDB Endow. 13, 7 (March 2020), 1134-1146. [Online]. Available: http://www.vldb.org/pvldb/vol13/p1134-ping.pdf
2. Uzi Cohen, Batya Kenig, **Haoyue Ping**, Benny Kimelfeld, and Julia Stoyanovich. 2018. A Query Engine for Probabilistic Preferences. In Proceedings of the 2018 International Conference on Management of Data - SIGMOD 18, ACM Press, Houston, TX, USA, 1509-1524. DOI:https://doi.org/10.1145/3183713.3196923
3. Batya Kenig, Lovro Ilijasic, **Haoyue Ping**, Benny Kimelfeld, and Julia Stoyanovich. 2018. Probabilistic Inference Over Repeated Insertion Models. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, 18971904.
4. Luke Rodriguez, Babak Salimi, **Haoyue Ping**, Julia Stoyanovich, and Bill Howe. 2018. MobilityMirror: Bias-Adjusted Transportation Datasets. In Big Social Data and Urban Computing - First Workshop, BiDU@VLDB 2018, Rio de Janeiro, Brazil, August 31, 2018, Revised Selected Papers (Communications in Computer and Information Science), Springer, 18-39. DOI:https://doi.org/10.1007/978-3-030-11238-7_2
5. Batya Kenig, Benny Kimelfeld, **Haoyue Ping**, and Julia Stoyanovich. 2017. Querying Probabilistic Preferences in Databases. In Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems - PODS 17, ACM Press, Chicago, Illinois, USA, 2136. DOI:https://doi.org/10.1145/3034786.3056111
6. Bill Howe, Julia Stoyanovich, **Haoyue Ping**, Bernease Herman, and Matt Gee. 2017. Synthetic Data for Social Good. In Proceedings of Data for Good Exchange (D4GX). arXiv: 1710.08874. [Online]. Available: http://arxiv.org/abs/1710.08874
7. **Haoyue Ping**, Julia Stoyanovich, and Bill Howe. 2017. DataSynthesizer: Privacy-Preserving Synthetic Datasets. In Proceedings of the 29th International Conference on Scientific and Statistical Database Management - SSDBM 17, ACM Press, Chicago, IL, USA, 15. DOI:https://doi.org/10.1145/3085504.3091117

8. Batya Kenig, Benny Kimelfeld, **Haoyue Ping**, and Julia Stoyanovich. 2017. A Database Framework for Probabilistic Preferences. In Proceedings of the 11th Alberto Mendelzon International Workshop on Foundations of Data Management and the Web, Montevideo, Uruguay, June 7-9, 2017. [Online]. Available: http://ceur-ws.org/Vol-1912/paper1.pdf
9. Julia Stoyanovich, Lovro Ilijasic, and **Haoyue Ping**. 2016. Workload-driven learning of mallows mixtures with pairwise preference data. In Proceedings of the 19th International Workshop on Web and Databases - WebDB 16, ACM Press, San Francisco, California, 16. DOI:https://doi.org/10.1145/2932194.2932202

Working Paper

11. **Haoyue Ping** and Julia Stoyanovich. 2021. Most Expected Winner: An Interpretation of Winners over Uncertain Voter Preferences. arXiv:2105.00082 [cs] (April 2021). [Online]. Available: http://arxiv.org/abs/2105.00082

Research Projects

Probabilistic preference databases

Query preferences of items and their properties over probabilistic databases. For example, given the preference distribution of movies, we can query the preferences over movie directors or movie genres. Here the director and genre are properties of the movies. I developed several exact and approximate solvers, and achieved the state-of-the-art inference results w.r.t. efficiency and accuracy.

- *Publications:* VLDB 2020 [1], SIGMOD 18 [2], AAAI 2018 [3], PODS 2017 [5]

Winner determination for uncertain votes

Determine winners in elections with uncertain voter preferences. Previous studies quantified candidate performance by the probabilities that a candidate would win the election. We follow the principle of score-based rules that high-scoring candidates should be favored, and propose a novel winner semantics, the Most-Expected-Winner. We establish hardness of winner determination under the new semantics, identify tractable cases, and develop efficient solvers.

Publication: arXiv [11]

Privacy preserving synthetic data generation

Generate statistically similar synthetic datasets from a sensitive dataset. For example, a hospital wants to collaborate with data scientists and needs to share patient data with them. But this hospital cannot leak the patient privacy without applying for regulatory approval first. This hospital needs a tool to share synthetic data with the data scientists that is statistically similar to the real data. It uses Bayesian Networks to model the correlations of attributes in sensitive datasets, and applies differential privacy to the learned distributions from the sensitive datasets.

Publication: SSDBM 2018 [7]

Software: DataSynthesizer (https://github.com/DataResponsibly/DataSynthesizer)

Preference modeling

Model user preferences using mixtures of Mallows models. The preferences in real world have many formats, including complete rankings, partial rankings, ratings, and pairwise preferences. This project regards pairwise preferences as the "building blocks" of arbitrary preferences, and models these preferences effectively with concise Mallows models. I made a contribution to a Java Library called *pref* for preference data mining, and developed the experiment code and visualized experimental results.

Publication: WebDB 2018 [9]

Academic Service

Program Committee Member

Extending Database Technology (EDBT) 2022

Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI) 2022

Teaching Experiences

Teaching Assistant for CS and EE courses at Drexel University

2018	CS	461	Database Systems
2015	ECEC	302	Digital System Projects
2015	ECEC	304	Design with Micro-controllers
2015	ECEC	355	Computer Organization & Architecture
2014	ECES	352	Introduction to DSP