

Haoyue Ping

PHD CANDIDATE, COMPUTER SCIENCE, NYU TANDON

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

RESEARCH INTERESTS

- Develop new techniques for preference modeling and inference
 - Build probabilistic databases for preference data management
 - Boost data sharing with Differential Privacy and Data Ethics
 - Solve real world problems with Data Science
-

EDUCATION

New York University *Jan' 2019 - Present*
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)

Drexel University *Sep' 2014 - Dec' 2018*
Ph.D. candidate 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

RESEARCH PROJECTS

Probabilistic-Preference-Inference

Supervisor : Dr. Julia Stoyanovich

Aug '16 - Present

- 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.
- Devised an algorithm to evaluate queries under Min-Max semantics.
- Implemented the inference algorithms of labeled RIM using Java.
- Published papers in PODS, SIGMOD, VLDB, and AAAI.

Data-Synthesizer

Supervisor : Dr. Julia Stoyanovich

Jan '16 - Present

- 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, which typically takes 18 months! This hospital needs a tool to share synthetic data with the data scientists which is statistically similar to the real data.
- Applied Bayesian Networks to model the correlations of attributes in sensitive datasets.
- Applied Differential Privacy to the learned distributions from the sensitive datasets.
- Open sourced this project [DataSynthesizer](#) on GitHub.
- Published a SSDBM demo paper.

Preference-Modeling

Supervisor : Dr. Julia Stoyanovich

Dec'15 - Jul' 16

- Model user preferences using mixtures of Mallows models. The preferences in real world has many formats including complete rankings, partial rankings, ratings, pairwise preferences, etc. This project regards pairwise preferences as the “building blocks” of arbitrary preferences, and model the preferences effectively with concise Mallows models.
- Pre-processed datasets into well-structured data and documented it.
- Contributed to a Java Library called *pref* for preference data mining.
- Developed the experiment code in Java and visualized experiment results in Python.
- Published a WebDB paper.

PUBLICATIONS

Supporting Hard Queries over Probabilistic Preferences,
Haoyue Ping, Julia Stoyanovich, and Benny Kimelfeld.
Proceedings of VLDB 2020.

A Query Engine for Probabilistic Preferences,
Uzi Cohen, Batya Kenig, **Haoyue Ping**, Benny Kimelfeld, Julia Stoyanovich.
Proceedings of ACM SIGMOD 2018.

Probabilistic inference over repeated insertion models,
Batya Kenig, Lovro Ilijasic, **Haoyue Ping**, Benny Kimelfeld, and Julia Stoyanovich.
Proceedings of AAAI 2018.

MobilityMirror: Bias-Adjusted Synthetic Transportation Datasets
Luke Rodriguez, Babak Salimi, **Haoyue Ping**, Julia Stoyanovich and Bill Howe
BiDU, 2018

Querying probabilistic preferences in databases,
Batya Kenig, Benny Kimelfeld, **Haoyue Ping** and Julia Stoyanovich,
Proceedings of PODS 2017.

Synthetic data for social good,
Bill Howe, Julia Stoyanovich and **Haoyue Ping**,
Proceedings of Data for Good Exchange (D4GX) 2017.

DataSynthesizer: Privacy-preserving synthetic datasets,
Haoyue Ping, Julia Stoyanovich and Bill Howe,
Proceedings of SSDBM 2017.

A database framework for probabilistic preferences,
Batya Kenig, Benny Kimelfeld, **Haoyue Ping** and Julia Stoyanovich,
Proceedings of AMW 2017.

Workload-driven learning of mallows mixtures with pairwise preference data,
Julia Stoyanovich, Lovro Ilijasic, and **Haoyue Ping**,
Proceedings of WebDB 2016.

TEACHING EXPERIENCE

Teaching Assistant for CS and EE courses in Drexel University

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