Jiwon Chang

+1 (585) 210-5512
jchang38@cs.rochester.edu
https://jiwonc.net

1701 South Avenue
Rochester, NY 14620

RESEARCH INTEREST

My current research interest is designing **(core)set query algorithms** for efficient and robust model training. I've worked on submodular quota sampling, gradient approximation and slice finding, among other objectives. In terms of technique, I frequently work with sampling algorithms to quickly maximize the objective(s) over big data and cope with unprocessed data in the wild. I believe randomized techniques will be integral for NVMe databases and interactive analysis.

EDUCATION

Ph.D. in Computer Science — Rochester, NY
University of Rochester — Aug 2023 – Present
Advisor: Fatemeh Nargesian
Relevant coursework: Data Management Systems, Machine Learning, Sampling Algorithms, Compilers

B.S. in Computer Science — Rochester, NY
University of Rochester — Aug 2018 – Dec 2022
Relevant coursework: Research Seminar, Data Mining, Artificial Intelligence, Data Structures & Algorithms

RESEARCH

Data Distribution Tailoring Revisited: Cost-Efficient Integration of Representative Data. Jiwon Chang, Bohan Cui, Fatemeh Nargesian, Abolfazl Asudeh, H.V. Jagadish. (Revision submitted, VLDBJ 2024.)

May 2022 – Jan 2023

We study the problem of cost-efficiently collecting a dataset that adequately represents discrete groups of interest in a data market setting. We are motivated by the scenario where additional data must be acquired to improve underperforming slices. Our journal paper drastically revamps algorithms and analysis.

- Designed and analyzed algorithms to efficiently satisfy quota sampling queries.
- Proposed a heuristic algorithm and a bandit algorithm with tight asymptotic guarantees.
- Developed an efficient linear algebra-based implementation of our algorithms.

Evaluating Performance Variance for Submodular Gradient-Approximating Coresets. Michael Flynn, Jiwon Chang, Zhengbin Tao, Fatemeh Nargesian. (Manuscript.)

Jul 2023 – Present

Submodular gradient approximating subsets are the current SOTA techniques for data-efficient machine learning. Yet, they are prone to high variance, which could cause group-wise accuracy disparity. We aim to better understand the behavior of gradient subsets in this aspect with a comprehensive empirical evaluation.

- Implemented group-fair variants of submodular gradient-approximating subsets in Go.
- Developed fairness metric for gradient-approximating subsets.

PLUTUS: Understanding Distribution Tailoring for Machine Learning. Jiwon Chang, Christina Dionysio, Fatemeh Nargesian, Matthias Boehm. (Under review, SIGMOD 2024.)

Dec 2023 – Jan 2023

PLUTUS is a system for interactive model debugging. We combine Sliceline, a sparse linear algebra algorithm for identifying problematic slices, Distribution Tailoring, a suite of algorithms for efficiently satisfying quota sampling queries from a union of heterogeneous sources, and an interactive dashboard. We demonstrate the usefulness of the human-in-the-loop system for navigating accuracy-fairness-cost tradeoffs.

- Implemented efficient quota sampling algorithms with Numpy and PostgreSQL.
- Developed an interactive dashboard for PLUTUS in Python using Plotly Dash.
- Created a demonstration video which walks through a Flights arrival delay case study.

ColorLang: A Language For Physically Correct and Computationally Efficient Color Programming. Ethan Chen, Jiwon Chang, Sreepathi Pai, Yuhao Zhu. (Under review, ASPLOS 2024.)

Jan 2023 – May 2023

Color science programs are prone to user error, frequently resulting in subtle physical or perceptual errors that are hard to debug. Furthermore, optimizing large tensor operations is often unintuitive. We design a tensorflow-like metaprogramming system for color science programming which enforces physical correctness with type checking and achieves high performance with tensor algebra optimization.

- Developed a metaprogramming language for developing physically correct color science programs.
- Implemented an equality saturation optimizer for tensor algebra execution plan.

GRANTS

Schwartz Discover Grant for Undergraduate Summer Research. Data Distribution Tailoring Revisited. $5000.

May 2022 - Aug 2022

PRESENTATIONS

UR Computer Science Open House. Data Distribution Tailoring Revisited.

Oct 2023

PROFESSIONAL EXPERIENCE

University of Rochester   Rochester, NY
Research assistant   Mar 2023 – May 2023
- Conducted computer science research at the Data Intelligence Lab and the Horizon Lab.
- Lead the Data Distribution Tailoring Revisited project.
- Assisted equality saturation implementation and paper writing for the ColorLang project.

BusySquirrels Company   Remote
Full-stack web developer intern   May 2021 – Aug 2021
- Improved BusySquirrels Shopping Tips browser extension's UI with React.js.
- Streamlined IndexedDB database queries for the browser extension.
- Implemented analytics across the product suite to track customer engagement.

Freelance   Remote
Private physics tutor   Aug 2018 – Feb 2022
- Taught the International Baccalaureate (IB) HL Physics curriculum.

ACTIVITIES

Campus Times   University of Rochester
Opinions Writer   Oct 2020 – Sep 2022

- Contributed opinion articles on technology and STEM education.

- Interviewed and reported CS undergraduates' sentiments on the curriculum.

PROFESSIONAL AFFILIATIONS

- Association of Computing Machinery (ACM)

- Special Interest Group in Management of Data (SIGMOD)

SKILLS

- **Programming languages:** Python, Javascript, Java, Go

- **Tools & Frameworks:** Tensorflow, pandas, scikit, React.js, git & Github, AWS, SystemDS

- **Databases:** MongoDB, PostgreSQL, DynamoDB, IndexedDB