Ziming Mao

421 Temple St, New Haven, CT 06511 | (203) 435-3725 | ziming.mao@yale.edu

RESEARCH INTERESTS

Big Data Systems, Machine Learning Systems, Cloud Computing, Distributed Systems

EDUCATION

Yale University, New Haven, CT 2019 - Expected Graduation 2023
Intended double major: B.S. in Computer Science, B.A. in Philosophy, GPA: 3.93 / 4.0

Hwa Chong Institution, Singapore 2015 – 2019

PUBLICATIONS

System:

- Juncheng Yang, **Ziming Mao**, Rashmi Vinayak, Yao Yue, "Learning at the level of object groups for high-efficiency high-throughput caching" Submitted to Networked Systems Design and Implementation (NSDI) 2023
- **Ziming Mao**, Anurag Khandelwal, Kiran Srinivasan, "*Trinity*: A Fast and Space-efficient Multi-attribute Data Store" Submitted to Operating Systems Design and Implementation (OSDI) 2022

NLP / Machine Learning

- **Ziming Mao***, Chen Henry Wu*, Ansong Ni, et. al. "*DYLE*: Dynamic Latent Extraction for Abstractive Long-Input Summarization" Accepted at Association for Computational Linguistics (ACL) 2022 Main Conference (* Equal Contribution)
- Yusen Zhang, Ansong Ni, **Ziming Mao**, et. al. "*Summ*^{N}: A Multi-Stage Summarization Framework for Long Input Dialogues and Documents" Accepted at Association for Computational Linguistics (ACL) 2022 Main Conference
- Linyong Nan, Chia-Hsuan Hsieh, **Ziming Mao**, et. al. "*FeTaQA*: Free-form Table Question Answering" Accepted at Transactions of the Association for Computational Linguistics (TACL) 2022
- Xiangru Tang, Alexander R. Fabbri, **Ziming Mao**, et. al. "Investigating Crowdsourcing Protocols for Evaluating the Factual Consistency of Summaries" Accepted at the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) 2022 Main Conference.

Earlier Works

- Yong Zen Tan, **Ziming Mao**, Yanjun Zhang, et. al. "Enhancing fouling mitigation of submerged flat-sheet membranes by vibrating 3D-spacers". Accepted at the journal of Separation and Purification Technology, 2019
- Bing Wu, Yanjun Zhang, **Ziming Mao**, et. al. "Spacer vibration for fouling control of submerged flat sheet membranes". Accepted at the journal of Separation and Purification Technology, 2019

RESEARCH

TheSys Group, Parallel Data Lab, CMU, *advised by Prof. Rashmi K. Vinayak* January 2022 - Current

- *GL-Cache*: group-level learning for cache eviction (with **Twitter**):
  - Incremental training and feature selection using XGBoost, and object merging metric for *GL-Cache*
  - Ported Cacheus, a fully adaptive ML caching algorithm with scan-resistant LRU and churn-resistant LFU as voters, from Python to C++
  - Implemented Segment-level LRU on cache simulator
  - Evaluations for *GL-Cache*, LRB, LHD, Cacheus, LeCaR, etc. on simulator

Computer Systems Lab, Yale, *advised by Prof. Anurag Khandelwal and Kiran Srinivasan* March 2021 – Current

- Virtualization for disaggregated architecture
  - *Ongoing*
- *Trinity*: space-efficient multi-attribute data store (with **NetApp**):
  - Designed *MdTrie*: a dynamic, succinct, self-indexed multi-dimensional data structure based on Morton encoding

- Proposed *Trinity*: a distributed sharded space-efficient storage system using Thrift
- Introduced additional storage optimizations: collapsed node representation and generalized Morton encoding
- Achieved at least 5x faster multi-attribute query and 2x lower storage compared to MySQL, MongoDB, and Aerospike on macrobenchmark and PH-Tree and R*-Tree on microbenchmark

LILY Lab, Yale, *advised by Prof. Dragomir Radev* October 2020 – Dec 2021

- Long-input Summarization (with **Microsoft Research**):
  - *DYLE*: Proposed and implemented the first summarization framework that jointly trains an extractor and a generator with extracted snippets as the latent variable; Introduced novel extractive oracles, consistency loss and hybrid training method.
  - *Summ^N*: A multi-stage framework that produces fine-grained summaries based on intermediate coarse summaries
  - Largely outperformed the state-of-the-art models on QMSum, GovReport, AMI, ICSI, SummScreen datasets.
- Table-based Question Answering: *FeTaQA* (with **Salesforce Research**)
  - Introduced a new dataset that includes tables, complex questions, and free-form answers pairs. Designed crowdsourcing protocols and organized annotations on Amazon MTurk
  - Contributed to a pipeline model (TaBERT + T5).
- Faithful Text Generation (with **Facebook AI**)
  - Investigating crowdsourcing protocols for evaluating factual consistency of summaries
  - Run state-of-the-art models (e.g. BART and Pegasus) on XSum and CNN/Dailymail datasets

Nanyang Technological University, *advised by Prof. Bing Wu* May 2017 – March 2018

- Proposed and patented 3D scraper-like spacer for membrane fouling control

SOFTWARE ENGINEERING

Altoroo, *Software Engineer Intern* Sept 2020 – January 2021

- Built a Python web scraper using Selenium and Beautiful Soup
- Incorporated GitHub "Single-Signed-On" into AWS Cognito using NodeJS OIDC wrapper

Cyabra, *Software Engineer Intern* June 2020 – August 2020

- Used Flask to generate social media scans. Implemented interactive analytics charts to display scan results
- Used ReactJS, Redux to build a platform from scratch for clients to monitor and allocate scan quotas

Yhack, *Tech Team Chair* February 2020 – May 2021

- Led a team of seven; Built Yhack and Yale Tech Conference website with ReactJS, Tailwind CSS, and Django
- Implemented mass-email, admin dashboard, team formation platform, typewriter animation, scroll-to-sections navbar

Yale x Flatiron Web Development Bootcamp May 2020 – July 2020

- Built a Ruby app that processes, plots and exports Covid-19 data of multiple countries
- Implemented a JavaScript platform game with user login and leaderboard using Ruby on Rails as the backend

COURSES

System: Operating Systems, Distributed Systems, Big Data Systems, Computer Networks, Computer Organization
ML/AI: Advanced NLP, Machine Learning, Computer Vision, Neural Networks of Linguistic Structure, Deep Learning
Theory/Math: Intensive Algorithm, Data Structures, Automata Theory, Discrete Math, Probability Theory

SKILLS

- *Programming:*
  Over 5k lines: C/C++, Python, Go, JavaScript, HTML, CSS, Latex
  Over 1k lines: ReactJS, Django, Ruby, Ruby on Rails
- *Languages:* Native fluency in English and Chinese
- *Interest:* Astrophotography, Badminton