BIJAN MAZAHERI

 github.com/honeybijan bijanmazaheri.com bmazaher@caltech.edu

 Pasadena, CA (781)-985-0881

I am a computer scientist and mathematician interested in the synthesis and transportability of knowledge from and between multiple data sources. My work spans the fields of causality, data fusion, and distribution shift. I am currently pursuing a Ph.D. in Computing and Mathematical Sciences at Caltech supported by a NSF Graduate Research Fellowship and an Amazon AI4Science Fellowship.

EDUCATION

California Institute of Technology - Pasadena, CA   *Oct 2017 - present*
Ph.D. Candidate
NSF Graduate Research Fellowship
Amazon AI4Science Research Fellowship

Cambridge University (Emmanuel College) - Cambridge, UK   *Oct 2016 - Jun 2017*
Mathematics Part 1B
Supported by a Herchel Smith Fellowship
Additional classes in Computer Science and Mathematics Part II

Williams College - Williamstown, MA   *Sep 2012 - Jun 2016*
Bachelor of Arts
Majors: Physics and Computer Science, GPA: 3.92
Honors: Highest Honors (Physics), Phi Beta Kappa, Sigma Xi, Magna Cum Laude

RESEARCH TOPICS

Data Fusion: Most datasets restrict their focus to a subset of topics. We have developed "Expert Graphs" to study consistencies and inconsistencies in partial, but overlapping expertise. As it turns out, this problem is deeply related to issues in voting theory, such as the Cordorcet Paradox.

Mixture Models for Causality: Combining datasets induces universal confounding on a structural equation model (SEM). Controlling for such a counfounder involves treating each dataset separately, defeating the purpose of the combined dataset. Assuming a bound on the number of types of sources turns the problem into a mixture model, allowing identification of within-source probability distributions. This perspective expands the notion of causal identifiability, as many graphically unidentifiable relationships can be identified.

Domain Adaptation and Transportability: I am interested in using causal principles to guide models that are robust to distribution shift in datasets. Our work uses concepts from causality and information theory to inform how auxiliary training tasks can improve robustness to distribution shift.

WORK EXPERIENCE

Amazon Research Causality Lab - Tübingen, Germany   *Oct 2022 - Feb 2023*
Applied Scientist Intern (L5)

Worked with Mila Hardt, Atalanti Mastakouri, and Dominik Janzing
Paper written and accepted to UAI 2023

BioDiscovery - El Segundo, CA *Jun 2017 - Sep 2017*
Intern
Unsupervised learning on cancer genomes. Developed and implemented methods that lead to a conference poster.

IBM T.J. Watson Research Center - Yorktown Heights, NY *Jun 2016 - Sep 2016*
Intern
Worked with Dr. Victor Kravets (mentor) and Dr. Andrew Sullivan (manager).
Projects included non-greedy and map-reduce algorithms for factoring sum of products

TEACHING EXPERIENCE

Markov Chain Monte Carlo *Spring 2022*
Head TA for Leonard Schulman's new class.

PUBLICATIONS

B. Mazaheri, A. Mastakouri, D. Janzing, M. Hardt. *"Causal Information Splitting: Engineering Proxy Features for Robustness to Distribution Shifts."* UAI 2023 (to appear).

S. Gordon, ***B. Mazaheri**, Y. Rabani, L. Schulman. *"Identifying Discrete Mixtures of Bayesian Networks."* CLeaR 2023.

S. Jain, **B. Mazaheri**, N. Raviv, J. Bruck. *"Glioblastoma signature in the DNA of blood-derived cells"* PLoS ONE 16(9): e0256831. 2021.

B. Mazaheri, S. Jain, J. Bruck. *"Expert Graphs: Synthesizing New Expertise via Collaboration."* IEEE ISIT 2021.

S. Gordon, ***B. Mazaheri**, Y. Rabani, L. Schulman. *"Source Identification for Mixtures of Product Distributions."* COLT 2021.

B. Mazaheri, S. Jain, J. Bruck. *"Robust Correction of Sampling Bias using Cumulative Distribution Functions."* NeurIPS 2020.

* = Authorship order is alphabetical.

PREPRINTS

B. Mazaheri. *"Evaluation of Causal Structure Contamination due to Dataset Merging."*

B. Mazaheri, S. Jain, M. Cook, J. Bruck. *"Combining Binary Classifiers Leads to Nontransitive Paradoxes."*
Accepted IEEE Transactions on Information Theory 2022 but retracted due to inability to attend conference in person.

S. Gordon, ***B. Mazaheri**, Y. Rabani, L. Schulman. *"The Sparse Hausdorff Moment Problem, with Applications to Topic Models."* arXiv 2020.

S. Jain, **B. Mazaheri**, N. Raviv, J. Bruck. *"Cancer Classification from Healthy DNA using Machine Learning."* bioRxiv 2019.

S. Jain, **B. Mazaheri**, N. Raviv, J. Bruck. *"Short Tandem Repeats Information in TCGA is Statistically Biased by Amplification."* bioRxiv 2019.

* = Authorship order in Theoretical CS is alphabetical. Co-authorship is implied.

PATENTS

S. Jain, **B. Mazaheri**, N. Raviv, J. Bruck. *"Mutation profile and related labeled genomic components, methods and systems."* 2019.

PROJECTS

LACCTiC.com   *Sep 2021 - present*
Website for ranking collegiate cross country with 10,000 regular users.
Algorithms/machine learning for adjusting times in different conditions.
Backend using Django, frontend using React, database hosted on AWS.

AWARDS

National Science Foundation Graduate Research Fellowship   *Awarded Spring 2019*
Awarded in 2019 for a proposal to research confounding influence in causal networks.
Amazon AI4Science Research Fellowship   *Awarded Spring 2022*
Funding for research with the potential to aid scientific discovery.

WORKSHOPS

Simon's Institute for Theory of Computing: Causality   *Spring 2022*
4 week workshop on Causal Inference methods.

TALKS

Simon's Institute for Theory of Computing: Causality Reunion   *May 2023*
Title: "Causal Discovery under Limited Global Confounding"

MENTORSHIP

Caltech Cross Country Team   *Sep 2018 - present*
Assistant Coach
Mentoring and supporting undergraduate students at Caltech.