



BIJAN MAZAHERI

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

 Pasadena, CA  (781)-985-0881

I am a computer scientist and mathematician interested in hidden variables in causal structures: when and if they can be learned and how they affect networks of opinions. While machine learning is often focused on building models from data, my work flips the narrative to explore what *models* can tell us about *incomplete or absent data*. I am currently pursuing a Ph.D. in Computing and Mathematical Sciences at Caltech supported by a NSF Graduate Research Fellowship.

EDUCATION

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

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

California Institute of Technology - Pasadena, CA

Oct 2017 - present

Ph.D. Candidate

NSF Graduate Research Fellowship

RESEARCH TOPICS

Expert Graphs: We are developing a structure to study consistency in overlapping expertise. This work helps us understand non-transitive properties in networks of machine learning classifiers as well as how to combine classifiers to synthesize new knowledge.

Source Identification: Controlled experiments and causal inference require an understanding of confounding variables. We use mixture models to learn unobservable confounding variables given assumptions about the visible variables. This allows us to perform causal inference without measuring, or even knowing about hidden confounders.

Covariate Shift: Preparing a machine learning classifier for demographic differences between training data and deployment. I am interested in developing stable and reliable methods for real world situations that diverge from the available training data.

WORK EXPERIENCE

IBM T.J. Watson Research Center - Yorktown Heights, NY

Jun 2016 - Sep 2016

Intern

Worked with Dr. Victor Kravats (mentor) and Dr. Andrew Sullivan (manager).

Projects included non-greedy and map-reduce algorithms for factoring sum of products

BioDiscovery - El Segundo, CA

Jun 2017 - Sep 2017

Intern

Unsupervised machine learning on cancer genomes. Developed and implemented methods that lead to a conference poster.

PUBLICATIONS

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. “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 in Theoretical CS is alphabetical. Co-authorship is implied.

PREPRINTS

S. Gordon, ***B. Mazaheri**, Y. Rabani, L. Schulman. ”Identifying Discrete Mixtures of Bayesian Networks.” Submitted to STOC 2022.

B. Mazaheri, S. Jain, J. Bruck. “Expert Graphs: Synthesizing New Expertise via Collaboration.” Submitted to IEEE Transactions on Information Theory.

B. Mazaheri. “A Global Entropic Measure of Latent Confounding Influence.” Submitted to AISTATS 2022.

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.

Algorithms/machine learning for adjusting times in different conditions.

Full stack development using SQL, Django, React, AWS.

Over 200,000 page views from 6,000 unique users in the first two months.

AWARDS

National Science Foundation Graduate Research Fellowship

Awarded Spring 2019

Awarded in 2019 for a proposal to research confounding influence in causal networks.

MENTORSHIP

Caltech Cross Country Team

Sep 2018 - present

Assistant Coach

Mentoring and supporting undergraduate students at Caltech.