David Arbour

Email:darbour26@gmail.com
Phone:(413)-559-8898
Website:darbour.github.io

Work Experience

11/2018 - Present **Adobe Research**
Research Scientist San Jose, CA

- Authored multiple publications in causal inference, policy evaluation and network analysis.
- Created a root-cause analysis model for diagnosing faults in distributed computing environments.
- Designed and deployed a model for measuring the effect of changes in data centers which accommodates irregular observations and multiple changes/interventions.
- Conducted causal analysis of product feature upgrades on user retention and usage rates for multiple Adobe products.
- Devised an implemented improvements to A/B testing framework by implementing regression adjustment with always valid confidence intervals.
- Deployed an online monitoring tool for detecting anomalies in computer system metrics.

10/2016 - **Facebook**
11/2018 - *Research Scientist* Menlo Park, CA

- Designed and deployed a model for video player and upload optimization.
- Contributed to a platform for automatic experimentation and optimization.
- Contributed to a library for performing Bayesian optimization.
- Developed methods for causal inference policy evaluation.

6/2011 - **UMass Amherst**
10/2016 *Research Assistant* Amherst,MA

- Conducted research investigating causal discovery in relational domains resulting in multiple publications.
- Designed machine learning pipeline for forecasting future engine part failures for Pratt & Whitney.
- Performed quasi-experimental designs (QEDs) to identify heterogeneity in the underlying process of publication and citation across scientific publication venues.
- Created models for citation prediction and anomaly detection using network analysis and machine learning for the IARPA SCITE project.

6/2014 - **Rocana (acquired by Splunk)**
3/2015 *Data Science Consultant* Remote

- Created and prototyped real-time anomaly detection algorithms for detecting unusual events in distributed systems.
- Researched and developed an automatic template discovery algorithm for machine generated log messages.

6/2013 - **Apple**
8/2013 *Data Science Intern* Cupertino, CA

- Developed and prototyped a novel temporally aware collaborative filtering algorithm for iOS app recommendation.
- Performed ad-hoc analyses of app store user behavior.

Education

2017 University of Massachusetts Amherst; Amherst, MA
Ph.D. Computer Science
Advisor: David Jensen
Thesis: "Methods for Enabling Causal Inference in Relational Domains"

2/2015 University of Massachusetts Amherst; Amherst, MA
M.S. Computer Science
Advisor: David Jensen

5/2010 University of Massachusetts Amherst; Amherst, MA
B.S. Computer Science
Cum Laude

Relevant Competencies

Languages
Python, R, SQL
Software/Environments
PyTorch, Stan, NumPyro, Git, Bash

Publications

Working Papers

- Ian Waudby-Smith, David Arbour, Ritwik Sinha, Edward H. Kennedy, and Aaditya Ramdas. Doubly robust confidence sequences for sequential causal inference, 2021

Conferences and Journals

- David Arbour, Drew Dimmery, and Arjun Sondhi. Permutation weighting. In *International Conference on Machine Learning (To Appear)*, 2021
- David Arbour, Drew Dimmery, and Anup Rao. Efficient balanced treatment assignments for experimentation. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, 2021
- My Phan, David Arbour, Drew Dimmery, and Anup Rao. Designing transportable experiments under s-admissability. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, 2021
- Ryan A. Rossi, Nesreen K. Ahmed, Aldo Carranza, David Arbour, Anup Rao, Sungchul Kim, and Eunyee Koh. Heterogeneous graphlets. In *Transactions on Knowledge Discovery from Data (TKDD)*, page 43, 2020
- Arjun Sondhi, David Arbour, and Drew Dimmery. Balanced off-policy evaluation in general action spaces. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, 2020
- Eli Sherman, David Arbour, and Ilya Shpitser. General identification of dynamic treatment regimes under interference. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, 2020
- David Arbour, Dan Garant, and David Jensen. Inferring network effects from observational data. In *Proceedings of the Twenty-Second Conference on Knowledge Discovery and Data Mining*, 2016
- David Arbour, Katerina Marazopoulou, and David Jensen. Inferring causal direction from relational data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, 2016
- Marc Maier, Katerina Marazopoulou, David Arbour, and David Jensen. A sound and complete algorithm for learning causal models from relational data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, 2013
- Paul E Dickson, David T Arbour, W Richards Adrion, and Amanda Gentzel. Evaluation of automatic classroom capture for computer science education. In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education*, 2010

- Paul E Dickson, W Richards Adrion, Allen R Hanson, and David T Arbour. First experiences with a classroom recording system. In *ACM SIGCSE Bulletin*, 2009

Workshops

- Ryan A Rossi, Nesreen K Ahmed, Aldo Carranza, David Arbour, Anup Rao, Sungchul Kim, and Eunyee Koh. Heterogeneous network motifs. *KDD 2019 Workshop on Machine Learning in Graphs*, 2019
- Marazopoulou Katerina, David Arbour, and David Jensen. On causal analysis for heterogeneous networks. The 2017 ACM SIGKDD Workshop on Causal Discovery, 2017
- David Arbour, Katerina Marazopoulou, and David Jensen. Look both ways: Dependence and direction in relational data. Workshop on Information in Networks, 2015
- Katerina Marazopoulou, David Arbour, and David Jensen. Refining the semantics of social influence. Networks: From Graphs to Rich Data, NIPS Workshops, 2014
- David Arbour, Katerina Marazopoulou, Dan Garant, and David Jensen. Propensity score matching for causal inference with relational data. Causal Inference: Learning and Prediction Workshop, UAI, 2014
- Marc Maier, Katerina Marazopoulou, David Arbour, and David Jensen. Flattening network data for causal discover: What could go wrong? Workshop on Information in Networks, 2013
- David Arbour, James Atwood, Ahmed El-Kishky, and David Jensen. Agglomerative clustering of bagged data using joint distributions. Structured Learning: Inferring Graphs from Structured and Unstructured Inputs Workshop, ICML, 2013
- Marc Maier, Katerina Marazopoulou, David Arbour, and David Jensen. A sound and complete algorithm for learning causal models from relational data. *Approaches to Causal Structure Learning Workshop, UAI*, 2013

Talks

- "Designing Transportable Experiments." Conference on Digital Experimentation (2020).
- "Permutation Weighting." UC Berkeley causal inference reading group (2020).
- "SoftBlock: Efficient and Optimal Treatment Assignment for Experiments." Conference on Digital Experimentation (2019).
- "Inferring network effects from observational data.". KDD (2016).
- "Causal Inference:From Simple Experiments to Relational Data". UMass Research Experience for Undergraduates (2016).
- "Causal Graphical Models of Relational Domains". Harvard Causal Inference Reading Group. (2016)
- "Causal Inference from Observational Relational Data". Facebook. (2016)
- "Look Both Ways: Dependence and Direction in Relational Data". 2015 Workshop on Information Networks.
- "Understanding Causality in Networks". UMass Research Experience for Undergraduates. (2014)
- "Relational Propensity Score Matching". 2014 UAI Workshop, Causal Inference: Learning and Prediction.

Professional Activities

Conference Program Committees:

- AAAI (2020–2021)
- AISTATS (2020–2021)
- ICML (2019–2021)
- ICWSM (2019–2021)
- ICLR (2019–2021)
- NeurIPS (2016, 2018–2021)
- UAI (2017–2021)
- WSDM (2021, Outstanding reviewer award)

- WWW (2017, 2018)

Journal Reviewing
European Journal of Operations Research (2016, 2020)
Management Science (2021)

Mentor for Ian Waudby-Smith (CMU, Adobe Research intrn 2020), Eli Sherman (JHU, Adobe Research intern 2019), Arjun Sondhi (UW, Facebook Research Intern 2018), Ahmed El-Kishky (UMass REU, 2013) and Molly McMahon (UMass Independent Study, 2015)