David Arbour

Email:darbour@cs.umass.edu
Phone:(413)-559-8898
Website:darbour.github.io

Research Interests

I develop novel methodologies using ideas from machine learning and social science to detect and measure causal dependence in relational datasets. I am also interested in the design of experiments and observational studies for relational domains.

Education

2016 (Expected) University of Massachusetts Amherst; Amherst, MA
Ph.D. Computer Science
Advisor: David Jensen
Thesis: "Measuring Causal Dependence in Relational Data"

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

Work Experience

Current – 6/2011 **UMass Amherst**
Research Assistant Amherst,MA
- Conducted research leveraging graphical models and statistical relational learning for improved automated causal discovery algorithms.
- Designed model for predicting future part failures for Pratt & Whitney, resulting in an order of magnitude increase in forecasting ability.
- Performed quasi-experimental designs (QEDs) to identify heterogeneity in the underlying process of publication and citation across scientific publication venues.

3/2015 – 6/2014 **Scaling Data (now Rocana)**
Data Science Consultant
- 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.

8/2013 – 6/2013 **Apple**
Data Science Intern Cupertino, CA
- Developed and prototyped a novel temporally aware collaborative filtering algorithm for iOS app recommendation.
- Performed ad-hoc analyses of user behavior of on the app-store.
- Created web-based analytics dashboard using Javascript and Python.

12/2010 **Vertica Systems (acquired by HP)**
6/2010 *Intern* Billerica, MA

- Developed web-based reporting system in Java for company-wide testing framework.
- Created tools for historical analysis of test-failures and in-depth reporting of individual errors.
- Designed database runtime reporting web application using Javascript and PHP.

6/2010 **UMass Amherst**
9/2008 *Research Assistant* Amherst, MA

- Rewrote video processing program an automatic lecture capture system, allowing for in process analysis and writing of video, reducing runtime from 8 hours to near real-time.
- Evaluated system by conducting surveys of students in courses using the lecture capture system.

8/2009 **IBM WebAhead Group**
6/2009 *Intern* Cambridge, MA

- Wrote intranet crawler to discover intra-company RSS feeds using Hadoop.
- Crafted analysis of RSS feed data, showing connections within company communications and impact of inner company blogs.
- Implemented email-based intra-company RSS subscription service using Python and Django; service was highest user rated product produced by a WebAhead intern.

Relevant Competencies

Languages
Python, R, SQL, Java, C

Software/Environments
Postgres, Hive, Impala, Git, Bash

Publications

- David Arbour, Katerina Marazopoulou, and David Jensen. Testing for dependence and direction in relational data. *In Submission*.
- David Arbour and David Jensen. Learning with mixtures of dependency networks. *In Submission*.
- 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. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, 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.
- 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*, pages 88–92. ACM, 2010.
- Paul E Dickson, W Richards Adrion, Allen R Hanson, and David T Arbour. First experiences with a classroom recording system. *ACM SIGCSE Bulletin*, 41(3):298–302, 2009.

Talks and Posters

- "Look Both Ways: Dependence and Direction in Relational Data". 2015 Workshop on Information Networks. (Talk)
- "Understanding Causality in Networks". 2015 UMass Research Experience for Undergraduates lunch. (Talk)
- "Relational Propensity Score Matching". 2014 UAI Workshop, Causal Inference: Learning and Prediction. (Talk)
- "Learning with Mixtures of Dependency Networks". 2014 New England Machine Learning Day. (Poster).
- "Agglomerative Clustering of Distributions". 2013 New England Machine Learning Day. (Poster)

Awards and Service

- Baystate Fellow
- Member Phi Kappa Phi
- Research experience for Undergradutes (REU) mentor for Ahmed El-Kishky (2013)