

A Software Suite for Causal Modeling and Discovery

Joseph Ramsey, PhD - Department of Philosophy, Carnegie Mellon University, Pittsburgh, Pennsylvania, PA, USA

Kevin Bui, BS - Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA;

Chirayu Wongchokprasitti, MS, PhD - Department of Biomedical Informatics, School of Medicine / School of Information Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Zhou Yuan - Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Mark Silvis, BS - Department of Computer Science, University of Pittsburgh, Pittsburgh, Pennsylvania, USA;

Michael Davis, MS - Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Jitske Venema Shunfenthal, BA, MA- Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Nick Nystrom, PhD - Pittsburgh Supercomputing Center / Carnegie Mellon University, Pittsburgh, PA, USA

Alexandros Labrinidis, PhD - Department of Computer Science, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Panos K Chrysanthis, PhD - Department of Computer Science, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Gregory Cooper, MD, PhD - Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

In the past 25 years, tremendous progress has been made in developing general computational methods for discovering causal knowledge from data based on a representation called causal Bayesian networks. While much progress has been made in the development of these computational methods, they have not been readily available, sufficiently efficient, or easy to use by biomedical scientists, and they have not been designed to exploit Big Data that are increasingly available for analysis.

The Center for Causal Discovery has created a suite of tools that make efficient causal modeling and discovery (CMD) algorithms from Big Data available on a variety of platforms and environments. The suite uses a common set of CMD algorithms implemented as a Java library. We have created software around this library to develop our suite:

Tetrad-lib – a readily imported Java library of CMD algorithms

Tetrad – a desktop application that runs on any Java-enabled computer

Causal-cmd – a command-line application that runs on any Java-enabled computer

Causal-web – an easy-to-use Web-based application that submits causal discovery jobs to an HPC (e.g., three terabyte 64 core nodes at the Pittsburgh Supercomputing Center (PSC) or Amazon EC2)

R-causal – an R library

Py-causal – a Python module

A Docker instance that contains a ready-to-run instance of R-Studio and the R-causal

Causal-REST-API – a RESTful API hosted at 1) the PSC that submits CMD jobs to the Bridges supercomputer and 2) Amazon EC2

All of our software is open source and licensed under the GNU GPL such that it can be modified and incorporated into other software. The software and documentation is freely available from www.ccd.pitt.edu, and the source code is available from github.com/cmu-phil/tetrad and github.com/bd2kccd.

Session to which submitted: Software, Analysis, & Methods Development