Seminar Course on Causation

Course Syllabus: Fall 2015
Course #80-816 (graduate level)
Course #80-516 (undergraduate level)
(both Section A)

Instructor: Kun Zhang
Email: *TBA*
Phone: TBA
Baker Hall, Room TBA
Carnegie Mellon University
Office Hours: 14:30 – 16:30 Tuesdays and 10:00 – 12:00 Wednesdays
Other times by appointment

Description

Causal connections are usually more interesting or helpful than purely associational information. This course is mainly concerned with systematic approaches to discovering causal connections from data in various scenarios and the question why causation plays an important role in science, i.e., how it is helpful in understanding, decision making, and prediction in complex environments.

We will study the difference between causal and non-causal systems and make an attempt to characterize a causal system. Apart from identification of causal effects, we will explore two causality-related areas. One is causal discovery, i.e., going beyond the observational data to the underlying causal information. It is well known that "correlation does not imply causality," but we will make this statement more precise by asking what information in the data and what assumptions enable us to discover causal information from purely observed data. This will cover constraint-based causal discovery, causal discovery based on structural equation models, causal discovery from time series, difficulties in practical causal discovery, causality in neuroscience, causality in biology, and causality in economics and finance. More importantly, we will have the opportunity to solve problems in various fields from a causal perspective: participants may bring any causal problems they are interested in, and we will work together to find potential solutions. The other is how to properly make use of causal information. This includes counterfactual reasoning, improving machine learning in light of causal knowledge, and forecasting in nonstationary unseen environments.

Overall, this course aims to provide fundamentals of causal discovery and inference, review emerging methods for causal discovery, report their applications, find practical causal problems in various fields, and work out potential solutions.

Course objectives

As an outcome of this course, participants are expected to
- get familiar with causality-related concepts and principles and emerging approaches to causal discovery from observational data,
- get to know the state-of-the-art of causality research in different disciplines,
- identify and raise causal problems in your respective fields, and work together to find potential solutions, and
- find and research novel causal problems in your own disciplines

Course Prerequisites

This course is accessible to students from across disciplines. As a prerequisite, students are expected to be familiar with elementary probability theory and elementary statistical estimation theory.

We especially welcome students from different departments. They will hopefully give a more complete picture of the causality research, raise causal problems of general interest, and inspire cross-disciplinary collaboration and novel ideas.

Class Schedule

Class meetings consist of lecture presentations of principles and methodologies for causal discovery and inference and student presentations and discussions of the problems of interest.

- Week 1 (Aug. 31): Lectures on causality-related concepts and principles (e.g., manipulation, directed acyclic graphs, causal effects, predicting the effects of manipulations, and structural equation models) and constraint-based causal discovery

- Week 2 (Sept. 14): Lectures on causal discovery based on structural equation models (e.g., linear and various nonlinear models) and causal discovery from time series (e.g., nonparametric approaches, Granger causality with instantaneous effects, causal discovery from subsampled data or in the presence of confounding time series)

- Week 3 (Sept. 21): Lectures on causal discovery with hidden variables, philosophical reflections on recent causal discovery methods, using causal knowledge to improve machine learning, and open questions

- Weeks 4 - 12 (9 weeks): Student presentations and discussions

- Weeks 13 - 14 (2 weeks, Dec. 7 & 14): Student short presentations to summarize your achievements on the presented problems

Topics to Present and Requirements

Participants are strongly encouraged to present your own causality-related problems or data sets. If you don't like to do so, you can select one from the following topics (reading materials will be available online) or find another one together with me by the end of the third week.

- Finding causal links between genes, traits, and disease (c.f. "Integrative analysis of 111 reference human epigenomes," available at http://www.nature.com/nature/journal/v518/n7539/ full/nature14248.html)
- Practical issues in causal discovery from time series
- Causal analysis in stock market
- Causality in climate analysis (e.g., prediction and understanding of El Niño)
- Causality and prediction in nonstationary environments (e.g., how to improve the performance of Google Flue Trends)
- Domain specific causal discovery (e.g., for fMRI, MEG, economic data, financial data, and climate data)
- Causal discovery and machine learning, especially semi-supervised learning multi-task learning
- Causal discovery and complexity measures

- Counterfactual approaches to causal inference
- Finding causal knowledge and using it for crime control
- Causality-based computational social science
- Towards "universal" causal discovery

Participants are expected to present the selected topics, make progress on those topics, and summarize the achievements. You may complete them *alone or in two-person groups*. We will expect two-person groups to go further on the chosen topics.

- *Presentations of your selected topics*: All participants will present the chosen problems, or reading materials, or data sets of interest in class during weeks 4 - 12
- *Final report* (limited to 10 pages): All participants will write a short report to summarize the problems, their potential solutions, and expected output, or your own understanding of the reading materials.
- *Final short presentation* (10 minutes): All participants will give a short presentation to briefly report your achievements on the presented problems.

Grading

20% participation, 40% presentation of interesting problems or assigned reading materials, raising questions, and active involvement in discussions, 25% final report, 15% presentation of the final report.

Important: We will adopt *discipline-specific* evaluation criteria for students from different disciplines (e.g., philosophy, machine learning, statistics, computer science, psychology, information systems, social and decision sciences, public policy, and biology).

Course Materials

Reading materials will be available online or distributed in class. In addition, we will refer to several chapters of the following two books frequently:

- Peter Spirtes, Clark Glymour, and Richard Scheines. 2000. Causation, prediction, and search, 2nd edition. MIT Press, Cambridge, Massachusetts

- Judea Pearl. 2009. Causality: Models, Reasoning and Inference, 2nd edition. Cambridge University Press, Cambridge.