Causal discovery for CDS practitioners

Diviyan Kalainathan, Olivier Goudet, Isabelle Guyon, Michele Sebag, Philippe Caillou, Paola Tubaro

Goal: Making available to the CDS community simple and practical tools of causal discovery.

Background: Uncovering cause-effect relationships is central in many aspects of everyday life in both highly industrialized and development countries: what affects our health, the economy, climate changes, world conflicts, and which actions have beneficial effects? Establishing causality is critical to guiding policy decisions in areas including medicine and pharmacology, epidemiology, climatology, agriculture, economy, sociology, law enforcement, and manufacturing. The need for assisting policy making while reducing the cost of experimentation and the availability of massive amounts of "observational" data prompted the proliferation of proposed causal discovery techniques, but it is fair to say that to this day, they have not been widely adopted by scientists and engineers.

Part of the problem is the **lack of simple methods**, which can be "safely" used by practitioners, without risking obtaining **flawed results by violating complex conditions of applicability**, which can often not even be tested. To address these shortcomings, our team has developed **novel practical methodologies** based on results of **past challenges** (http://www.causality.inf.ethz.ch/challenge.php). Additionally, causal model offer new possibilities to address problems of transfer learning and missing data, which require an understanding of changes in distributions due to particular interventions on the system (see recent publications http://webdav.tuebingen.mpg.de/causality/).

Proposal: Our proposal addresses both education/outreach and research goals. From the **education/outreach** point of view, we want to **illustrate the range of applicability of causal modeling methods**. We intend to put together a benchmark suite, which can serve both as reference and as educational tool. We will wrap the sample tasks in **Jupyter notebooks** and upload them in **RAMP studio**. From the **research** point of view, we want to put our new methodology to work on **real world applications**. Such applications include projects we are already involved in:

(1) **A social science study about happiness in the workplace** based on two datasets : Dares survey (French ministry of labour) on 33000 individuals over 471 questions, and social and financial yearly reports of 2000 French companies since 2002 (SECAFI databases). Preliminary analyses based on correlations show promising results. Causal algorithms would offer new possibilities to make recommendations to managers based on the results provided by our analysis.

(2) **An application to improve the operation of the French power network** of "Réseau de Transport d'Électricité" (RTE). Determining causes of failures and corrective actions to protect the network from operating outside of safety is an ever challenging problem with new intermittent generation sources coming from "renewable energies" and the development of electricity trading between countries, putting pressure on a network with limited possibility of expansion.

(3) **An application in epidemiology.** In France, an estimated 180,000 people have type 1 diabetes (including 5,000 to 10,000 diabetic children. The causes of the disease are still unclear. We have available data from a cohort of approximately 6000 patients from a multicenter study conducted by INSERM (http://www.isis-diab.org/). It is a long and ambitious study gathering thousands environmental factors, survey answers, clinical data, genetic and epigenetic data.

While these three applications offer great material for our causal studies, we are totally open to accepting new propositions from other CDS members.

Funding:

- 2 months of a CDS engineer to assist us with using the RAMP platform.
- 3 master student internships, each to implement one case study with Jupyter notebook and RAMP implementation.