An integral approach to modeling causality

In this dissertation we tackle the problem of modeling causal relationships The complete task exists of: (a) choosing, (b) learning and (c) using these causal models. The choice of the model will describe the set of systems that can be represented and the types of questions that can be answered. Learning is what makes the models usable in practice.

We focus on probabilistic graphical models as they provide an intuitive and powerful representation. The causal interpretation of probabilistic graphical models has opened a wide area of research in recent years. Most of the work has focused on (a) causal inference, this is the calculation of the effect of an intervention on the system on some set of variables, (b) the elucidation of causal relationships from observational data and (c) formalizing more expressive causal models.

We identified some gaps in these three subareas and have proposed solutions toward forming an integral modeling approach of causal relationships.

In general learning from observational data alone yields a partial causal model, in order to construct the entire model experiments are needed. We propose an algorithm that allows to complete the structure by performing experiments. The goal is to minimize the total cost/number of experiments needed We then extend this approach for settings in which the provided data is flawed.

Unmeasured confounders, which are variables that influence the system but are not included in the model, complicate the learning process and complicate probabilistic/causal inference. We provide a learning algorithm that allows to learn models in this setting and provide a parametrization that helps to perform inference.

Most studies assume that a system is modeled in a centralized way. However, systems tend to get more and more complex and are sometimes inherently distributed. We propose a distributed extension of the causal models and provide learning and inference algorithms for this setting.

Stijn Meganck