Phase transition in detecting causal relationships from obervational and interventional data

Alexander Hartmann[*1] and Gregory Nuel[2]

[1]Institut of Physics, University of Oldenburg, – Carl-von-Ossietzky Straße 9-11, 26111 Oldenburg, Germany

[2]Laboratoire de Probabilités et Modèles Aléatoires (LPMA) – Université Pierre et Marie Curie (UPMC) - Paris VI, Université Pierre et Marie Curie [UPMC] - Paris VI – France

Abstract

Analysing data of, e.g., gene-expression experiments, and modelling it via network-based approaches is one of the main data analysis tasks in modern science. A first step is to use the observation of correlation to infer the structure of the underlying network, for example in the context of the inverse Ising model. Although this is already a algorithmically hard problem, it becomes even more challenging, when one is even interested in causal relationships between the network entities. When just using correlations obtained from observational data, such causal relationships can only be resolved partially.

One way out is to include interventions to the system, e.g., by knocking out genes when studying gene expression. This allows, in principle, to get a grip on the causal structure of a system. Here, we model the data using Gaussian Bayesian networks defined on directed acyclic graphs (DAGs). By applying an approach which allows for multiple interventions in each single experiment and which takes large-scale interaction effects into account by calculating joint maximum likelihoods (MLs) for rather large (sub) networks, causal relationships can be detected, in principle, with high accuracy. The algorithm which achieves this needs on top of the ML calculation to sample different causal orderings, which induce different DAGs. We present a new Monte Carlo approach to sample orderings, which is based on approximating the full ML by probabilities of orderings of triplets. We show that this approximation is rather good and efficient. This allows us to study the quality of the causality detection as a function of the fraction of interventional experiments. We observe a (information) phase transition between phases where the causal structure cannot be detected and where it can be detected. The transition point occurs roughly where only one intervention per network node is necessary.

*Speaker