EGU2020-9269
https://doi.org/10.5194/egusphere-egu2020-9269
EGU General Assembly 2020

Comparing Causal Discovery Methods using Synthetic and Real Data

Christoph Käding and Jakob Runge
German Aerospace Center (DLR), Institute of Data Science, Jena, Germany

Unveiling causal structures, i.e., distinguishing cause from effect, from observational data plays a key role in climate science as well as in other fields like medicine or economics. Hence, a number of approaches has been developed to approach this. Recent decades have seen methods like Granger causality or causal network learning algorithms, which are, however, not generally applicable in every scenario. When given two variables X and Y, it is still a challenging problem to decide whether X causes Y, or Y causes X. Recently, there has been progress in the framework of structural causal models, which enable the discovery of causal relationships by making use of functional dependencies (e.g., only linear) and noise models (e.g., only non-Gaussian noise). However, each of them is coming with its own requirements and constraints. While the corresponding conditions are usually unknown in real scenarios, it is quite hard to choose the right method for every application in general.

The goal of this work is to evaluate and to compare a number of state-of-the-art techniques in a joint benchmark. To do so, we employ synthetic data, where we can control for the dataset conditions precisely, and hence can give detailed reasoning about the resulting performance of the individual methods given their underlying assumptions. Further, we utilize real-world data to shed light on their capabilities in actual applications in a comparative manner. We concentrate on the case considering two uni-variate variables due to the large number of possible application scenarios. A profound study, comparing even the latest developments, is, to the best of our knowledge, so far not available in the literature.