Inferring predictive causal networks from multiple time-series

Rui Castro,
Department of Mathematics and Computer Science,
Eindhoven University of Technology

May 31, 2021

This project explores the construction of models for very large and complex systems based using data-driven approaches. Many systems are so large that they defy human-intuition. For example this is the case in biological systems where genes regulate the production of other genes, giving rise to very complex networks. Even for man-made devices, such as large machines, similar considerations apply. Even if the machines are designed by humans, the various components often interact in ways that are unexpected and defy intuition. This project is aims to explore how such models can be constructed in a meaningful and statistically sound way from observational data.

Our goal is to identify a *predictive causal network*, based on multiple time-series data. This model consists of a directed (and possibly) weighted simple graph consisting of n vertices/nodes. Each vertex in the graph corresponds to a device or sensor, endowed with a series of measurements. Specifically, let $t \in \{1, \ldots, t\} \equiv [T]$ denote time, where $T \in \mathbb{N}$ is the overall time horizon. Associated with each vertex $i \in [n]$ there is a time series $(X_{it},\ t \in [T])$. These time-series are related somehow, although we believe that, for many pairs of vertices there is a conditionally independence relation. Dependence relations are, on the other hand, directional (and this is a distinctive feature in comparison with many types of graphical models). So, in our predictive causal network a directed edge from vertex i to j indicates that the time series indexed by i is "better" at predicting the time-series indexed by j than the other way around. Overall we believe the underlying predictive causal network to be sparse, meaning the total number of edges in the network is relatively small (i.e., much smaller than $\binom{n}{2}$). The main research challenge is to devise a sound methodology to effectively learn such a network from multiple time-series data. This entails devising a suitable (but general enough) formulation that will lead to reasonable inference procedures.

At the heart of this challenge is the notion of predictive causality between a pair of time-series. Let $(A_t,\ t \in [T])$ and $(B_t,\ t \in [T])$ be two time-series. We say A_t predictive-causes B_t if the past of A_t and B_t is better at predicting the future of B_t, then what we could do if using the past of B_t alone. This concept was pioneered by Clive Granger and is known as Granger-causality. The classical formulations are in the context of autoregressive models, but there are more recent and general formulations using information theoretical notions, in particular that of *transfer entropy*. The first part of the project will explore the various possibilities to measure the "strength" of such predictive causal relations based on pairwise comparisons.

Regardless of which underlying methodology, ultimately the methods above will give a weight W_{ij} for the strength of the relation between i and j (these need to be interpreted with care). Note that in general $W_{ij} \neq W_{ji}$. The second part of the project is to investigate how to properly use these metrics to construct a sparse predictive causal network. Note that, unless the time horizon T is very large, it will not be possible to simultaneously determine the exact degree of relation for all pairwise relations. Therefore, there is only hope if one believes there is an underlying parsimony in such networks. How can we capitalize on this? In settings were undirected relations are present (measured through correlations) there are some sensible ways to proceed, such as the graphical Lasso and related methods. The main research question here is how can similar ideas be applied in a setting where instead of correlations one has directed relations. In a nutshell, while focusing on predictive causal relations, what are sensible ways to develop regularization procedures for the inference of these networks that are both sound (statistically speaking) and practical (algorithmically speaking). With a more practical mindset, a very concrete question is whether cross-validation and bootstrap approaches can also be used to determin the "right" amount of regularization.