
Spectral structure learning for clinical time series

Ivan Lerner^{*1}

¹Health data- and model- driven Knowledge Acquisition – Inria Paris, Institut National de la Santé et de la Recherche Médicale - INSERM – France

Résumé

Structure learning algorithm aims at learning the dependencies between variables, often represented in a directed acyclic graph (DAG), and which under strong additional assumptions can be interpreted as causal relations. Structure learning algorithms have been adapted to time series, from constraint-based algorithms, such as PC-MCI, to score-based algorithms, such as DYNOTEARS. Clinical time series are multivariate time series observed in multiple patients, which are irregularly sampled, and sampled at different frequencies, thus challenging existing structure learning algorithms. With this study, we aimed to develop and evaluate a structure learning algorithm for clinical time series.

We assume that our times series are realizations of StructGP, a k -dimensional multi-output or multi-task stationary Gaussian process (GP), with independent patients sharing the same covariance function. The covariance function is built in the framework of process convolution, by defining the observed process as the convolution between a k by k matrix-valued impulse response function and k -independent white noise processes. Each element of the impulse response function is defined as a squared-exponential. We further assume that the support of the impulse response function encodes a sparse DAG, through imposing the output scales of the squared-exponentials that form a k by k matrix to be sparse and lower-triangular up to permutations. This parametrization corresponds to having the frequency composition of the time series follows a linear additive Gaussian structural causal model at each frequency, with a structure common to all frequencies. We show that this DAG is identifiable and that it encodes ordered conditional relations between time series. We implement an adapted NOTEARS algorithm, which based on a differentiable definition of acyclicity, recovers the graph by solving a series of continuous optimization problems. The algorithm performance is evaluated by simulating random graphs of different degrees and measuring the capacity to recover the true graph. Finally, it is illustrated in a cohort of patients with Sickle cell disease.

Our results show that up to mean degree 3 and 20 tasks, most errors are spurious links. With 20 tasks, we reach a median recall of 0.93% (IQR, 0.86-0.97) while keeping a median precision of 0.71% (0.57-0.84) for recovering directed edges. We further show that the regularization path is key to identifying the graph.

With StructGP, we proposed a model of time series dependencies, that flexibly adapt to different time series regularity, while enabling us to learn these dependencies from observations. Further work will be needed to bridge the gap between performance in simulated data and real-world data, regarding the sensitivity to scaling and unmeasured confounders.

*Intervenant