

Gene networks from time series with spike-and-slab feature selection

Edgar Steiger, Martin Vingron

*Max Planck Institute for Molecular Genetics, Berlin, Germany;
Computational Systems Biology, Humboldt-Universität zu Berlin, Berlin, Germany*

The availability of time series gene expression data allows unravelling of gene regulatory networks with causal relations by the simple premise of "cause precedes effect". These effects do not take place immediately, but rather with a certain delay. The proposed algorithm is based on regression with feature selection to find the causal relations and the correct delays from given time series data. It does so in a two-fold way: it uses spike and slab priors in a Bayesian setting to put weight on the most important variables in the regression, and it imposes a multinomial distribution on groups of variables to choose the most important member of each group. These groups are defined inherently by the temporal organization of the data. The parameters are fitted to the model using the Expectation Propagation algorithm which takes advantage of the factorized representation of the posterior distribution. It should be mentioned that the algorithm is not restricted to time series data but can also be applied to other scenarios where variables are grouped and between-group and in-group sparsity are assumed, thus being an alternative to other sparse regression approaches. The algorithm is tested and compared to some other established approaches on available data with a "gold standard" network.