Poster J-40

Comparative Evaluation of the Accuracy of Reverse Engineering Gene Regulatory Networks with various Machine Learning Methods

Authors:
Adriano V. Werhli (*BioSS*)
Marco Grzegorczyk (*University of Dortmund*)
Dirk Husmeier (*BioSS*)
Wolfgang Urfer (*University of Dortmund*)

Short Abstract: We compare the accuracy of predicting gene regulatory networks with three different machine learning methods: (1) relevance networks, (2) graphical Gaussian models, and (3) Bayesian networks. The evaluation is carried out on a cellular signalling network that describes the interaction of 11 phosphorylated proteins and phospholipids in human immune system cells.

Long Abstract:

INTRODUCTION

An important problem in systems biology is to infer the architecture of biochemical pathways and regulatory networks from postgenomic data. Various reverse engineering methods have been proposed in the literature, and it is important to understand their relative merits and shortcomings. To shed light onto this problem, the present paper evaluates and compares the performance of different machine learning methods on real and simulated data.

METHODS

We compared three widely-used methodologies in our evaluation study: relevance networks (RNs), graphical Gaussian models (GGMs), and Bayesian networks (BNs). The method of RNs, proposed by Butte and Kohane (2003), is based on pairwise association scores between the nodes. This approach is straightforward to implement and not particularly computationally expensive in its execution. The principled disadvantage of RNs, however, is that the inference of an interaction between two nodes is not done in the context of the whole system, and the method cannot distinguish between direct and indirect associations. This shortcoming is addressed by GGMs, where we compute the partial correlation between two nodes, conditional on all the other nodes in the system. On the assumption of Gaussian-distributed data this allows us to distinguish between direct and indirect interactions. BNs are more flexible probabilistic graphical models for conditional dependence and independence relations. As opposed to RNs and GGMs, these graphs are directed, which can be exploited in interventional studies for identifying putative causal interactions. In

our study, we applied the shrinkage estimator of Schaefer and Strimmer (2005) to compute the inverse covariance matrix for GGMs. We sampled BNs from the posterior distribution with Markov chain Monte Carlo (MCMC), sampling over node orders, as proposed by Friedman and Koller (2003).

DATA

We based the evaluation of the reverse engineering methods on the protein signalling network reported in Sachs et al. (2005); this is a cellular signalling network that describes the interaction of eleven phosphorylated proteins and phospholipids in human immune system cells. We used four types of data for the evaluation: the measured protein activities reported in Sachs et al. (2005), and synthetically generated data. The latter were obtained from a modified steady-state approximation to an ordinary differential equation description of chemical kinetics. Each data type was further subdivided into observational and interventional data. Observational data are measurements obtained by passively monitoring the biological system without any interference. Interventional data are obtained by actively manipulating certain domain variables, e.g. using gene knock-outs or overexpressions. These values are, thus, no longer dependent on the intrinsic dynamics and stochasticity of the system, and this leads to important clues for the determination of the direction of causal interactions.

EVALUATION

The true network used in our evaluation is a directed graph. The inference methods applied to learning this network may lead to undirected, directed, or partially directed graphs. To assess the performance of these methods, we applied two different criteria. The first approach, referred to as the undirected graph evaluation (UGE), discards the information about the edge directions altogether. To this end, the original and the learned networks are replaced by their skeletons, where the skeleton of a network is defined as the network in which two nodes are connected by an undirected edge whenever these nodes are connected by any type of edge in the original network. The second approach, referred to as the directed graph evaluation (DGE), compares the predicted network with the original directed graph, where a predicted undirected edge is interpreted as a superposition of two directed edges, pointing in opposite directions.

Applying a learning algorithm to any of the methods included in our evaluation study leads to a matrix of scores associated with the edges in the network, which defines a ranking of the edges. From this ranking we can obtain the receiver operator characteristics (ROC) curve, where the relative number of true positive (TP) edges is plotted against the relative number of false positive (FP) edges. We pursued two different evaluation procedures. The first approach is based on integrating the ROC curve so as to obtain the area under the curve (AUC), with larger scores indicating, overall, a better performance. While this approach does

not require us to commit ourselves to the adoption of any (arbitrary) decision criterion, it does not lead to a specific network prediction. It also ignores the fact that, in practice, one is particularly interested in the performance for low FP rates. Our second performance criterion, hence, is based on the selection of a (low) threshold on the edge scores, from which a specific network prediction is obtained. This threshold is chosen such that the different methods achieve the same FP score. The resulting procedure is guaranteed to compare the competing methods at the same operation point on the ROC curve, and the evaluation can therefore simply be based on the TP counts.

RESULTS

Detailed results will be presented on the poster, but the main findings can be summarized as follows. On Gaussian observational data, BNs and GGMs were found to outperform RNs. The difference in performance was not significant for the non-linear simulated data and the cytoflow data, though. Also, we did not observe a significant difference between BNs and GGMs on observational data in general. However, for interventional data, BNs clearly outperformed GGMs and RNs, especially when taking the edge directions (DGE score) rather than just the skeletons of the graphs (UGE score) into account. This suggests that the higher computational costs of inference with BNs over GGMs and RNs are not justified when using only passive observations on a system, but that active interventions in the form of gene knockouts and over-expressions are required to exploit the full potential of BNs

---
REFERENCES

Butte and Kohane (2003) In: Parmigiani et al." "The Analysis of Gene Expression Data" Springer, 428-446

Friedman and Koller (2003) Machine Learning 50, 95-126

Sachs et al. (2005), Science 308, 523-529

Schafer and Strimmer (2005) Statistical Applications in Genetics and Molecular Biology 4 (1) Article 32