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

Adriano V. Werhli[1], *Marco Grzegorczyk*[2] *and Dirk Husmeier*[1]

1) Biomathematics & Statistics Scotland, Edinburgh EH9 3JZ, United Kingdom
2) Department of Statistics, University of Dortmund, Germany

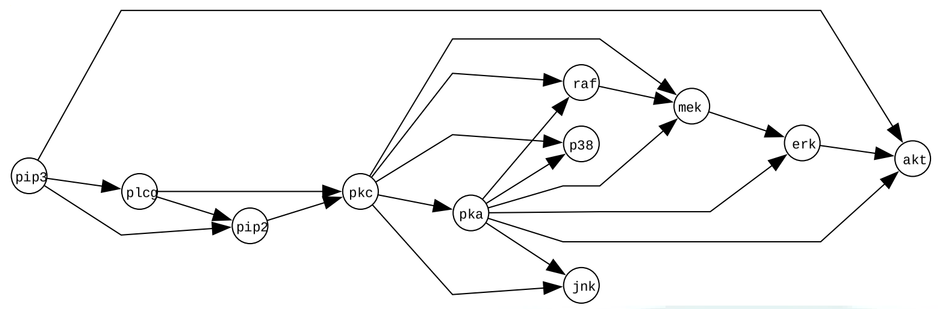
Abstract: We compare the accuracy of predicting gene regulatory networks with three different machine learning methods: (1) relevance networks, (2) graphical Gaussian models, and (3) Bayesian networks. The evaluation is carried out on a cellular signalling network that describes the interaction of 11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

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.

Method. We compared three widely-used methodologies in our evaluation study:

- **Relevance networks (RNs)** (Butte and Kohane, 2003): Interactions between genes are inferred from pairwise association scores. Disadvantage: Gene interactions are not inferred within the context of the whole system.
- **Graphical Gaussian models (GGMs)** (Schäfer and Strimmer, 2005): Gene interactions are computed from pairwise partial correlations, that is, the pairwise correlation between two nodes conditional on all the other nodes in the system.
- **Bayesian networks (BNs)** (Pearl, 1988): 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 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.



Raf signalling pathway. The graph shows the currently accepted signalling network, taken from Sachs et al. Nodes represent proteins, edges represent interactions, and arrows indicate the direction of signal transduction. In the interventional studies, the following nodes were targeted. Activations: PKA and PKC. Inhibitions: PIP2, AKT, PKC and MEK.

We used three types of data for the evaluation:

- **Synthetic data** from a Gaussian distribution.
- **Realistic simulated data** from a steady-state approximation to an ordinary differential equation description of chemical kinetics, using Netbuilder (Yuh et al., 1998).
- **Cytometry protein activities** reported in Sachs et al. (2005).

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.

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:

- **UGE:** undirected graph evaluation
- **DGE:** directed graph evaluation

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:

- **AUC:** Area under the curve, with larger areas indicating, overall, a better performance.
- **TP count:** True positive number of edges for the same false positive count of FP=5 across all methods.

Results 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.

References

- Butte and Kohane (2003), The Analysis of Gene Expression Data, Springer, 428-446
- Friedman and Koller (2003), Machine Learning 50, 95-126
- Pearl (1988), Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann
- Sachs et al. (2005), Science 308, 523-529
- Schäfer and Strimmer (2005), Bioinformatics 21, 754-764
- Yuh et al. (1998), Science 279, 1896-1902