LETTERS IN BIOMATHEMATICS
(URL to be inserted upon acceptance)

Strategies to avoid overfitting of MCMC Bayesian learning in some biological applications

Diego Garcia[a], Irene Tischer[a]

[a]School of Systems and Computing Engineering, Universidad del Valle, Santiago de Cali, Colombia

ABSTRACT

Model learning from observed data is typically affected by overfitting, because in order to find the models best parameter set, all relations between data are used indifferently whether they represent relevant or noisy interactions. Bayesian networks are widely used in biological modeling (e.g. networks of gene interactions), given that they allow representing graphically and determining statistically the dependence /independence relations between considered variables. A frequent approach in Bayesian learning is Markov Chain Monte Carlo simulation (MCMC), where a set of viable networks are explored by a random walk which converges to a network fitted optimally to data with respect to the likelihood or similar evaluation function. Here we propose various strategies to mitigate overfitting in Bayesian learning by MCMC in order to reduce the resulting models' complexity. They either apply constraints inside the MCMC simulation or consider post-optimal operations. We show the effectiveness of these strategies in some biological applications.

ARTICLE TYPE
Research Article

ARTICLE HISTORY
To Be Determined
To Be Determined

KEYWORDS
Bayesian networks, Bayesian learning, MCMC simulation, overfitting

1 Introduction

Bayesian networks are a powerful tool of knowledge representation and reasoning under uncertainty conditions, that often are present in real world applications. A Bayesian network is a directed acyclic graph, in which nodes represent random variables and edges denote dependencies between them.

There are three approaches for Bayesian learning when structure network is unknown: first approach is learning constraint-based, where Bayesian networks are seen as a representation of dependencies. In approach score-based, Bayesian networks are treated as a specification of a statistic model and then Bayesian learning is addressed to problem of model selection. In third approach instead to learn only one structure, it generate a set of feasible structures. This methods increase Bayesian reasoning, and try to average the prediction for each structure that belong to the set of possible structures.

Bayesian learning based in Markov Chain MonteCarlo (MCMC) typically works by simulating a Markov chain over the space of feasible networks structures, whose stationary distribution is the posterior distribution of the network.

2 Materials and methods

Sectioning commands work just as they do in the `article` document class.

2.1 Materials

If you followed how sectioning commands work, then you might have guessed how to get a subsection.

CONTACT Diego M. Garcia `diego.mauricio.garcia@correounivalle.edu.co`

2.2 Methods

In fact, this document class was built using the `article` document class. Hopefully, if you started with that document class or something analogous to it, converting to this document class is not too difficult.

3 Results and Discussion

Getting the hang of it yet?

3.1 Without strategy

3.2 Ilustration of strategies

3.2.1 Strategy One

3.2.2 Strategy Two

3.2.3 Strategy Three

Conclussions

Annex