Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

Signal Processing 86 (2006) 814–834

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks

Harri Lähdesmäki[a,b,*], Sampsa Hautaniemi[a], Ilya Shmulevich[c], Olli Yli-Harja[a]

[a]*Institute of Signal Processing, Tampere University of Technology, P.O. Box 553, FIN-33101 Tampere, Finland*
[b]*Cancer Genomics Laboratory, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA*
[c]*Institute for Systems Biology, Seattle, WA 98103, USA*

Received 26 January 2004; received in revised form 1 April 2005; accepted 14 June 2005
Available online 18 August 2005

Abstract

A significant amount of attention has recently been focused on modeling of gene regulatory networks. Two frequently used large-scale modeling frameworks are Bayesian networks (BNs) and Boolean networks, the latter one being a special case of its recent stochastic extension, probabilistic Boolean networks (PBNs). PBN is a promising model class that generalizes the standard rule-based interactions of Boolean networks into the stochastic setting. Dynamic Bayesian networks (DBNs) is a general and versatile model class that is able to represent complex temporal stochastic processes and has also been proposed as a model for gene regulatory systems. In this paper, we concentrate on these two model classes and demonstrate that PBNs and a certain subclass of DBNs can represent the same joint probability distribution over their common variables. The major benefit of introducing the relationships between the models is that it opens up the possibility of applying the standard tools of DBNs to PBNs and vice versa. Hence, the standard learning tools of DBNs can be applied in the context of PBNs, and the inference methods give a natural way of handling the missing values in PBNs which are often present in gene expression measurements. Conversely, the tools for controlling the stationary behavior of the networks, tools for projecting networks onto sub-networks, and efficient learning schemes can be used for DBNs. In other words, the introduced relationships between the models extend the collection of analysis tools for both model classes.

Keywords: Gene regulatory networks; Probabilistic Boolean networks; Dynamic Bayesian networks

*Corresponding author. Institute of Signal Processing, Tampere University of Technology, P.O. Box 553, FIN-33101 Tampere, Finland. Tel.: +358 3 3115 3859; fax: +358 3 3115 3817.

E-mail addresses: harri.lahdesmaki@tut.fi (H. Lähdesmäki), sampsa.hautaniemi@tut.fi (S. Hautaniemi), is@ieee.org (I. Shmulevich), yliharja@cs.tut.fi (O. Yli-Harja).

0165-1684/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2005.06.008

1. Introduction

During recent years, it has become evident that cellular processes are executed in a highly parallel and integrated fashion and that computational modeling approaches can provide powerful methodologies for gaining deeper insight into the operation of living cells. The modeling problem that has received a considerable amount of attention is the discovery of transcriptional level interactions. With the help of recent development in high-throughput genomic technologies, computational methods have enormous potential in the context of model inference from real measurement data and in practical use, such as drug discovery.

A number of different frameworks for gene regulatory network modeling have been proposed, ranging from differential equations too qualitative models (for an overview, see e.g. [1]). There is a clear conceptual difference between differential equation and coarse-scale models. The former can be used for a detailed representation of biochemical reactions, whereas the latter emphasize fundamental, generic principles between interacting components. In this context, models classes that are both discrete-time and discrete-state are called coarse-scale models.

Fine-scale modeling of biological interactions at the molecular level may require some type of differential equations. Although differential equations have successfully been used to simulate small (known) biochemical pathways (see e.g. [2,3]), their use in large-scale (genome-wide) modeling has considerable limitations. First of all, those models are computationally very demanding. Therefore, when modeling regulatory networks with differential equations, the model selection problem is usually ignored and the underlying biological system is assumed to be known. Because the model selection is the most important computational tool for discovering new, unknown regulatory relationships from the measured data, researchers have considered alternative modeling approaches. Also, the available analysis tools for differential equations are much more restricted than the ones for the alternative model classes (see below).

So-called graphical models can overcome the above-mentioned modeling problems, and advanced analysis tools have been developed for them. The use of holistic, coarse-scale models is also supported by the fact that the currently available data is limited both in quality and the number of samples. That is, there is no advantage using models that are much more accurate than the available data. Another constraint to be kept in mind is that the modeling framework should also be selected on the basis of the preferred goals, i.e., to what kinds of questions are we seeking answers. The two most often used large-scale modeling frameworks are Boolean and Bayesian networks (BNs). Since the Boolean network is a special case of another commonly used model class, probabilistic Boolean networks (PBNs), we will consider PBNs instead of Boolean networks.

PBNs is a model class that has been recently introduced in the context of genetic network modeling [4]. PBN is a stochastic extension of the standard Boolean network that incorporates rule-based dependencies between variables but is also stochastic in nature. The PBN model has a strong biological motivation through the standard, often used Boolean network model, originally proposed by Kauffman [5,6]. The theory of PBNs as models of genetic regulatory networks has been developed further in several papers. In particular, there has been interest in the control of stationary behavior of the network by means of gene interventions/perturbations [7], modifications of the network structure [8], and external control [9]. Another recent paper [10] introduces mappings between PBNs, including projections, node adjunctions and resolution reductions, which at the same time preserve consistency with the original probabilistic structure. Further, learning methods for PBNs have been introduced in [11,12]. More efficient learning schemes, in terms of computational complexity, but with cost of decreased accuracy, have been studied in [13]. General learning concepts have also been introduced in [14], although not in the context of PBNs, but a related setting. Kim and co-authors also show that the Markovian gene regulatory network model[1] is biologically plausible [15].

[1]The dynamics of PBNs can be studied using Markov chains.