

Construction of Probabilistic Boolean Networks from a Prescribed Transition Probability Matrix: A Maximum Entropy Rate Approach

Xi Chen¹, Wai-Ki Ching^{*,1}, Xiao-Shan Chen², Yang Cong¹ and Nam-Kiu Tsing¹

¹ *Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong.*

² *School of Mathematical Sciences, South China Normal University, Guangzhou, China.*

Received 8 March 2010; Accepted (in revised version) 20 September 2010

Available online 7 April 2011

Abstract. Modeling genetic regulatory networks is an important problem in genomic research. Boolean Networks (BNs) and their extensions Probabilistic Boolean Networks (PBNs) have been proposed for modeling genetic regulatory interactions. In a PBN, its steady-state distribution gives very important information about the long-run behavior of the whole network. However, one is also interested in system synthesis which requires the construction of networks. The inverse problem is ill-posed and challenging, as there may be many networks or no network having the given properties, and the size of the problem is huge. The construction of PBNs from a given transition-probability matrix and a given set of BNs is an inverse problem of huge size. We propose a maximum entropy approach for the above problem. Newton's method in conjunction with the Conjugate Gradient (CG) method is then applied to solving the inverse problem. We investigate the convergence rate of the proposed method. Numerical examples are also given to demonstrate the effectiveness of our proposed method.

Key words: Boolean networks, conjugate gradient method, genetic regulatory networks, inverse problem, Markov chains, Newton's method, probabilistic Boolean networks, transition-probability matrix.

1. Introduction

Building mathematical models and developing efficient numerical algorithms for studying regulatory interactions among DNA, RNA, proteins, and small molecules are important research issues in computational systems biology [7, 26]. In fact, many formalisms and

*Corresponding author. *Email addresses:* dlkcissy@hotmail.com (X. Chen), wching@hkusua.hku.hk (W.-K. Ching), cxs333@21cn.com (X.-S. Chen), congyang0305@yahoo.com.cn (Y. Cong), nktsing@hku.hk (N.-K. Tsing)

mathematical models have been proposed in the literature to study genetic regulatory networks such as Bayesian networks [25], Boolean Networks (BNs) [21, 22], multivariate Markov chain models [9], regression models [45], Probabilistic Boolean Networks (PBNs) [32–35], and a review on other mathematical models can also be found in [16, 36]. Among these models, BNs and their extensions PBNs have received much attention as they are able to capture the switching behavior of biological processes [26].

Boolean logic owes its name to George Boole who devised a mathematical framework for logical reasoning [4, 5]. BN models were first introduced by Kauffman [21–24]. Reviews of BN models can be found in [26, 37]. In a BN, the gene expression states are quantized to only two levels: on and off (represented as 1 and 0). The target gene is determined by several genes called its input genes via a Boolean function. When the input genes and the Boolean functions are given, then we say that a BN is defined. We remark that a BN is a deterministic model and the only randomness comes from its initial state. Given an initial state, the BN will eventually enter into a cycle of states called its attractor cycle. Since genetic regulation processes exhibit uncertainty and microarray data sets used to infer the model have errors due to experimental noise in the complex measurement processes, it is more realistic to consider stochastic models. The idea of extending the concept of a BN (a deterministic model) to a PBN (a probabilistic model) is as follows. For each gene, there can be more than one Boolean function and corresponding selection probabilities are assigned to the Boolean functions. The dynamics (transitions) of a PBN can be studied using Markov chain theory [10, 32, 35].

Given a PBN, the network behavior is characterized by its steady-state probability distribution which gives the first-order statistical information of a PBN. One can understand a genetic regulatory network and identify the influence of different genes via such a network. In [44], an efficient method has been used to construct the transition probability matrix and the standard iterative power method for computing the resulting steady-state probability distribution. Later, also a matrix approximation method has been proposed in [11] to get an approximation of the steady-state probability distribution efficiently. Furthermore, it is possible to control some genes in a network so as to drive the whole network into a desirable state or a steady-state probability distribution (a mixture of states). Therapeutic gene intervention or gene control policy [12, 15, 33, 35] can therefore be developed and studied.

Here we study the problem of constructing a PBN based on a given transition-probability matrix and a set of BNs. This is an inverse problem of huge size. The inverse problem is ill-posed, meaning that there can be many networks or no network having the desirable properties. Pal et al. [29] have presented two algorithms to solve the inverse problem of finding attractors constituting a BN. Network inference from steady-state data is a very important problem as most microarray data sets are assumed to be obtained from sampling the steady-state. In fact, the inverse problem can be split into two different tasks. The first task is to construct a sparse transition-probability matrix from a given network steady-state probability distribution. A maximum entropy rate approach has been proposed for this purpose [13]. The second task is to construct a PBN (the BNs and the selection probabilities) from a given steady-state probability distribution. Here, we propose to apply Newton's