The Fourth International Conference on Computational Systems Biology (ISB2010)
Suzhou, China, September 9–11, 2010

A Modified Entropy Approach for Construction of Probabilistic Boolean Networks

Xi Chen[1] Limin Li[1] Wai-Ki Ching[1] Nam-Kiu Tsing[1]

[1]Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Hong Kong.
Emails: dlkcissy@hotmail.com, { liminli, wching}@hkusua.hku.hk, nktsing@hku.hk

Abstract Boolean Network (BN) and its extension Probabilistic Boolean network (PBN) have received much attention in modeling genetic regulatory networks. In this paper, we consider the problem of constructing a PBN from a given positive stationary distribution. The problem can be divided into two subproblems: Construction of a PBN from a given sparse transition probability matrix and construction of a sparse transition matrix from a given stationary distribution. These are inverse problems of huge sizes and we proposed mathematical models based on entropy theory. To obtain a sparse solution, we consider a new objective function having an addition term of L_α-norm. Newton's method in conjunction with CG method is then applied to solve the inverse problem. Numerical examples are given to demonstrate the effectiveness of our proposed method.

Keywords Genetic Regulatory Networks, Sparse Probabilistic Boolean Networks; Inverse Problem; L_α-morm.

1 Introduction

The study of mathematical models and efficient numerical algorithms for the regulatory interactions among DNA, RNA, proteins and small molecules is an important research topic in genomic research [9]. In fact, a lot of formalisms have been proposed to study genetic regulatory networks such as Bayesian networks [13], Boolean Networks (BNs) [11], multivariate Markov chain model [3], Probabilistic Boolean Networks (PBNs) [16, 17]. A Reviews on mathematical models can also be found in [8]. Among these models, BN and its extension PBN have received more and more attention as they are able to capture the switching behavior of biological processes [9]. BN model was first introduced by Kauffman [11, 12]. Interested readers can find reviews of BN models in [9]. In a BN model, each gene is regarded as a vertex in the network and the gene expression states are quantized to only two levels: on and off (represented as 1 and 0). The target gene is predicted by several genes called its input genes via a Boolean function. We say a BN is defined if the input genes and the Boolean functions are given. We remark that a BN is actually a deterministic model and the only randomness comes from its initial state. Considering the inherent deterministic directionality in BNs as well as only a finite number of possible states, it is directly to see that some states will be re-visited infinitely, depending on the initial starting state. Such states are called *attractors* and the states lead to them comprise their *basins of attraction*. The number of transitions needed to return to a given

state in an attractor is called the cycle length [11]. It is also well known that eventually a BN will enter into an attractor cycle and stay there forever. In fact, the cycles can have biological significance [10] such as states of cell proliferation or cell apoptosis. For more details we refer interested readers to [12].

Since genetic regulation process exhibits 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 a stochastic model, Probabilistic Boolean Network (PBN). The idea of extending the concept of a BN (a deterministic model) to a PBN is as follows. For each gene, there can be more than one Boolean function (a set of Boolean functions with selection probabilities assigned to them). The dynamics (transitions) of a PBN can be described using Markov chain theory [4, 16, 17]. Given a PBN, assuming that the underlying Markov chain is irreducible, the long-run behavior is characterized by its stationary distribution which gives the first-order statistical information of a PBN. One can understand a genetic network and also study the influence of different genes via such a network.

Here we address the problem of constructing a PBN from a given stationary probability distribution. This problem is very important to network inference from steady-state data, as most microarray data sets are assumed to be obtained from sampling the steady-state. For the case of BN, Pal, et al. [14] have proposed two algorithms to solve the inverse problem of finding attractors constituting a BN. For the case of PBN, some preliminary works have been done base on entropy theory [7, 18]. Motivated by the results in [7, 18], we tackle the inverse problem by splitting into two different inverse problems. One of the problem (Problem 1) is to construct a PBN (the BNs and the selection probabilities) from a given sparse transition probability matrix. Newton's methods in conjunction with CG method has been proposed in [2] to solve the problem. Another problem (Problem 2) is to construct a sparse transition probability matrix from a given positive stationary distribution. An entropy rate approach has been proposed for this purpose [6]. To get more sparse solution in both problems, here consider adding a term in L_α-norm [1] to the objective function.

The remainder of the paper is structured as follows. Section 2 gives a mathematical formation of the inverse problems. Section 3 gives some numerical examples to demonstrate our proposed methods. Finally, concluding remarks are given in section 4 to address further research issues.

2 The Inverse Problems

In this section, we discuss the mathematical formulation of Problems 1 & 2. We first begin with Problem 1, i.e. to construct a PBN from a given sparse transition probability matrix with a set of BNs. We then proceed with Problem 2, to construct a spares transition probability matrix from a given positive stationary distribution.

2.1 A New Mathematical Formulation for Problem 1

Given a transition probability matrix A and a set of Boolean networks $\{A_i\}$ (biological rules), we are to construct a PBN. Since the problem size is huge and A is usually very sparse, here we assume that each column of A has m non-zero entries. In this case, we have $N = m^{2^n}$ and we can order $A_1, A_2, \cdots, A_{m^{2^n}}$ systematically. We note that q_i and A_i are

non-negative and there are only $m \cdot 2^n$ non-zero entries in A. Thus we have $m \cdot 2^n$ equations for m^{2^n} unknowns. In other words, we are interested in the estimation of the parameters $q_i, i = 1,2,\ldots,m^{2^n}$ when A is given.

Then, one possible way to get q_i is to consider the following minimization problem ([2]):

$$\min_{\mathbf{q}} \left\| A - \sum_{i=1}^{m^{2^n}} q_i A_i \right\|_F^2 \tag{1}$$

subject to

$$0 \le q_i \le 1 \quad \text{and} \quad \sum_{i=1}^{m^{2^n}} q_i = 1.$$

Here q_i is the selection probability (importance) of the ith BN and $\|\cdot\|_F$ is the Frobenius norm of a matrix. For the given matrix $A = (a_{ij})_{l\times l}$, we can re-order the entries of A by defining a mapping function F from the set of $l \times l$ square matrices to the set of $l^2 \times 1$ vectors as follows:

$$F\left(\begin{pmatrix} a_{11} & \cdots & a_{1l} \\ \vdots & \vdots & \vdots \\ a_{l1} & \cdots & a_{ll} \end{pmatrix}\right) = (a_{11},\ldots,a_{l1},a_{12},\ldots,a_{l2},\ldots,\ldots,a_{1l},\ldots a_{ll})^T. \tag{2}$$

If we let

$$U = [F(A_1),F(A_2),\ldots,F(A_{m^{2^n}})] \quad \text{and} \quad \mathbf{p} = F(A) \tag{3}$$

then (1) becomes

$$\min_{\mathbf{q}} \|U\mathbf{q} - \mathbf{p}\|_2^2 \tag{4}$$

subject to

$$0 \le q_i \le 1 \quad \text{and} \quad \sum_{i=1}^{m^{2^n}} q_i = 1.$$

In practice, we note that the matrix U is usually very large, it is not possible to store the whole matrix and therefore we need to seek for iterative methods for solving the above minimization problem. One possible and reasonable approach is to consider the solution which has the largest entropy rate as $\mathbf{q}$ itself can be considered as a probability distribution. This means we would like to find $\mathbf{q}$ such that it maximizes

$$-\sum_{i=1}^{m^{2^n}} q_i \log(q_i). \tag{5}$$

Entropy can be considered as a measurement of the uncertainty associated with a random variable [15]. It measures, in the sense of an expected value, the information contained in a message. Entropy method can also be regarded as a measure of the multiplicity associated with the state of the objects. Suppose we are given a state which can be accomplished in much more ways. Then it is more preferable than one which can be accomplished in only a few ways.

2.1.1 A Modified Entropy Approach

In [1], the authors present a algorithm for reconstructing a sparse solution $\mathbf{x} = (x_1, \ldots, x_n)$ from a small number of constraints by solving a linear system. They demonstrated that, by adding 1-norm of $\mathbf{x}$, i.e. $\sum_{i=1}^{n} |x_i|$, to the objective function, it is more likely to get a sparse solution. Motivated by this idea, we would like to modify the objective function in 5 However, in our problem, we have the constraint that

$$\sum_{i=1}^{m^{2^n}} q_i = 1 \quad i = 1, \ldots, m^{2^n}$$

and this means $\sum_{i=1}^{m^{2^n}} |q_i| = 1$ and the L_1-norm actually has no effect. Hence, we need to modify the norm by the following

$$\beta \sum_{i=1}^{m^{2^n}} q_i^{\alpha} \tag{6}$$

where α and β are two parameters. We adopt both the variance (the larger the better) and entropy (the smaller the better) of $\mathbf{q}$ as two possible measurements of solutions (they give consistent results in our numerical experiment). In practice, we try different values of α and β to obtain the best result. We employ grid search approach for finding the optimal values of α and β with grid size of 0.01. Here α ranges from 0.01 to 0.99 and β ranges from 0.1 to 2.0.

By adding the extra term of the form (6), one can get a modified entropy approach as follows:

$$\max_{\mathbf{q}} \left\{ -\sum_{i=1}^{m^{2^n}} q_i \log q_i - \beta \sum_{i=1}^{m^{2^n}} q_i^{\alpha} \right\} \tag{7}$$

where $0 < \alpha < 1$ and $\beta \geq 0$. The first term is the entropy rate term as in (5) and the second term is the L_α-norm part which helps in getting a sparse solution $\mathbf{q}$. The new optimization problem can be formulated as follows:

$$\max_{\mathbf{q}} \left\{ -\sum_{i=1}^{m^{2^n}} q_i \log q_i - \beta \sum_{i=1}^{m^{2^n}} q_i^{\alpha} \right\} \quad \text{subject to} \quad \bar{U}\mathbf{q} = \bar{\mathbf{p}} \quad \text{and} \quad 0 \leq q_i \quad i = 1, 2, \ldots, m^{2^n}. \tag{8}$$

We then follow similar analysis in [2] and apply the Lagrange multiplier method to the optimization problem (8). To build the Lagrange function, we only involve the consideration of the constraint $\bar{U}\mathbf{q} = \bar{\mathbf{p}}$. The constraint $q_i \geq 0$ is checked in the whole process. Let $\mathbf{y}$ is the multiplier and $L(.,.)$ is the Lagrangian function, then we have

$$L(q, y) = \max_{\mathbf{q}} \left\{ -\sum_{i=1}^{m^{2^n}} q_i \log q_i - \beta \sum_{i=1}^{m^{2^n}} q_i^{\alpha} + \mathbf{y}^T (\bar{\mathbf{p}} - \bar{U}\mathbf{q}) \right\} \tag{9}$$

The optimization problem can then be solved by using Newton's method in conjunction with CG method [6].

2.2 A New Mathematical Formulation for Problem 2

For Problem 2, given a positive stationary distribution π of a finite Markov chain, say N states, we are construct a transition probability matrix P corresponding to it, i.e.,

$$P\pi = \pi \quad \text{and} \quad (1,1,\ldots,1)P = (1,1,\ldots,1). \tag{10}$$

It is clear that there can be infinite many possible solution for the captured problem. In [6], it was proposed to apply the generalized entropy rate as a measure of uncertainty (randomness) of the objective function:

$$\sum_{j=1}^{N} w_j \left(-\sum_{i=1}^{N} P_{ij} \log P_{ij} \right) \tag{11}$$

where

$$0 \le w_j \le 1 \quad \text{and} \quad \sum_{i=1}^{N} w_i = 1.$$

The parameter w_j represents the weighting (importance) of State j. We note that $-\sum_{i=1}^{N} P_{ij} \log P_{ij}$ is the entropy of the conditional probability distribution when the chain is in State j.

The authors in [6] proposed to use the generalized entropy rate with a penalty term as the objective function:

$$\sum_{j=1}^{N} w_j \left(-\sum_{i=1}^{N} P_{ij} \log P_{ij} \right) + \sum_{j=1}^{N} \theta_j \left(\sum_{i=1}^{N} (P_{ij} - \frac{1}{N})^2 \right) \tag{12}$$

subject to (10). The following parameters are pre-determined:

$$0 \le w_j \le 1, \quad \sum_{i=1}^{N} w_i = 1, \quad \text{and} \quad 0 \le \theta_j \le 1, \quad \sum_{i=1}^{N} \theta_i = 1.$$

We note that the first term is the entropy rate and the second term is a penalty cost. The penalty cost measures the deviation of conditional distribution at State j from the uniform distribution. The parameters w_i and θ_i are the weightings. We note that the first term is concave and the second term is convex. Therefore they introduce some extra conditions on θ_i and w_i so that the above maximization problem have a unique solution.

2.2.1 A Modified Entropy Rate Approach

Since here we want to construct transition probability matrix from the given stationary distribution π, we need to modify the objective function. Then the new optimization problem is the following:

$$\max_{P_{ij}} \left\{ \sum_{j=1}^{N} \pi_j \left(-\sum_{i=1}^{N} P_{ij} \log P_{ij} \right) - \sum_{j=1}^{N} \left(\beta \sum_{i=1}^{N} P_{ij}^{\alpha} \right) \right\} \tag{13}$$

subject to

$$\begin{cases} \sum_{i=1}^{N} P_{ij} & = & 1, j = 1,2,\cdots,N \\ P\pi & = & \pi \\ P_{ij} & \ge & 0, i,j = 1,2,\cdots,N. \end{cases} \tag{14}$$

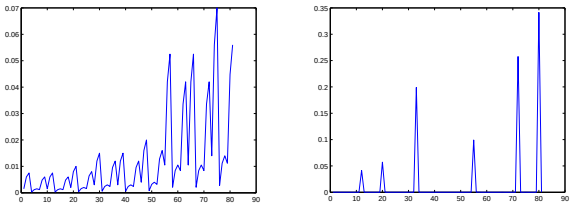

Figure 1: The Probability Distribution **q** for the Case of $A_{2,3}$. Method in [2] (Left) and Our method (Right)

The parameter w_j represents the weighting of state j.

Again we adopt both the weighted variances (the larger the better) and weighted entropies (the smaller the better) of the column distributions of P as two possible measurements of solutions (they give consistent results in our numerical experiments). The weightings are the probability mass of the given stationary distribution π. In practice, we try different values of α and β to obtain the best result. We also employ grid search approach for finding the optimal values of α and β with grid size of 0.01. Here α ranges from 0.01 to 0.99 and β ranges from 0.1 to 2.0.

3 Numerical Experiments

We first present two numerical examples of PBNs with $n = 2, m = 2$ and $n = 2, m = 3$ to demonstrate the proposed method for Problem 1. We then give an example used in [6] to demonstrate the proposed method for Problem 2.

3.1 Numerical Examples for Problem 1

In this section, we give two two-gene examples to compare with the method proposed in [2]. Using our new entropy approach, we obtain the solution as shown in Figure 1 (Right). The optimal solution is reached when $\alpha = 0.77$ and $\beta = 1.50$ in both cases. From the solution, we note that the re-constructed PBN is supposed to be dominated (over 99.7%) by the 6th, the 8th the 10th, the 12th, the 13th and the 15th BNs. From the dominated BNs, one can therefore construct the underlying regulatory rules, i.e., their truth tables. Here we see that our method can be used to identify the major components of the BNs constituting the PBN better than the method in [2], see Figure 1 (Left).

We then consider the case $n = 2$ and $m = 3$. The observed transition probability matrix of the PBN is given as follows:

$$A_{2,3} = \begin{pmatrix} 0.1 & 0.3 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.6 & 0.4 \\ 0.7 & 0.4 & 0.0 & 0.5 \end{pmatrix}. \tag{15}$$

Using our modified entropy approach, the optimal solution is reached when $\alpha = 0.61$ and $\beta = 0.6$. It is clear that our method give a "sparser" solution.

3.2 A Numerical Example for Problem 2

In this section, we give a numerical example for Problem 2 to compare with the solution obtained by the method proposed in [6]. Given the stationary distribution $\pi = (0.1, 0.2, 0.3, 0.4)$ of a Markov chain of four states, we want to construct the transition probability matrix corresponding to it. In [6], the optimal solution obtained is given as follows:

$$P_1 = \begin{pmatrix} 0.1860 & 0.1344 & 0.0947 & 0.0653 \\ 0.2390 & 0.2220 & 0.2010 & 0.1784 \\ 0.2741 & 0.2918 & 0.3031 & 0.3083 \\ 0.3009 & 0.3518 & 0.4012 & 0.4480 \end{pmatrix}.$$

Using our our method, we get the optimal transition probability matrix as follows:

$$P_2 = \begin{pmatrix} 0.0000 & 0.0830 & 0.1126 & 0.1240 \\ 0.0000 & 0.2097 & 0.2234 & 0.2276 \\ 0.1902 & 0.3250 & 0.3115 & 0.3063 \\ 0.8098 & 0.3824 & 0.3525 & 0.3420 \end{pmatrix}.$$

The optimal solution is reached when $\alpha = 0.94$ and $\beta = 1.6$. It is clear the our method gives a "sparser" solution.

4 Concluding Remarks

We present two modified entropy methods for constructing a PBN from a given sparse transition probability matrix and constructing a sparse transition probability from a given stationary distribution. Newton's method in conjunction with CG method is then applied to solving the inverse problem. We also give the sparsity comparison with other methods. The preliminary results of some small size networks is encouraging. There are two major limitations of our proposed method. The size of the problem increases exponentially with respect to the number of genes, in further research, we will consider much larger size examples and designing suitable preconditioners so as to accelerate the convergence rate of the CG method. Moreover, the study of the optimal parameters α and β is another interesting issue.

Acknowledges

Research supported in part by HKRGC Grant No. 7017/07P, HKUCRGC Grants, HKU Strategy Research Theme fund on Computational Sciences, National Natural Science Foundation of China Grant No. 10971075 and Guangdong Provincial Natural Science Grant No. 9151063101000021.

References

[1] E. J. Candes and T. Tao (2006), Near-optimal signal recovery from random projections: universal encoding strategies. IEEE Trans. Inform. Theory, 52 5406-5425.

[2] X. Chen, W. Ching, X. Chen, Y. Cong, N. Tsing (2010), Construction of Probabilistic Boolean Networks from a Prescribed Transition Probability Matrix: A Maximum Entropy Rate Approach, submitted.

[3] W. Ching, E. Fung, M. Ng and T. Akutsu, (2005), On Construction of Stochastic Genetic Networks Based on Gene Expression Sequences, International Journal of Neural Systems, (15), 297-310.

[4] W. Ching and M. Ng (2006) Markov Chains : Models, Algorithms and Applications, International Series on Operations Research and Management Science, Springer: New York.

[5] W. Ching, S. Zhang, Y. Jiao, T. Akutsu and A. Wong, (2009), Optimal Control Policy for Probabilistic Boolean Networks with Hard Constraints. IET on Systems Biology, (3), 90-99.

[6] W. Ching and Y. Cong, (2009), *A New Optimization Model for the Construction of Markov Chains*, CSO2009, Hainan, IEEE Computer Society Proceedings, 551-555.

[7] W. Ching, X. Chen and N. Tsing, (2009), Generating Probabilistic Boolean Networks from a Prescribed Transition Probability Matrix, IET on Systems Biology, 6, 453-464.

[8] H. de Jong, (2002), Modeling and Simulation of Genetic Regulatory Systems: A Literature Review, J. Comput. Biol., (9), 69-103.

[9] S. Huang, (1999) Gene Expression Profiling, Genetic Networks, and Cellular States: An Integrating Concept for Tumorigenesis and Drug Discovery, J. Mol. Med., (77), 469-480.

[10] S. Huang and D.E. Ingber, (2000), Shape-dependent Control of Cell Growth, Differentiation, and Apoptosis: Switching Between Attractors in Cell Regulatory Networks, Exp. Cell Res., (261), 91-103.

[11] S. Kauffman, (1969), Metabolic Stability and Epigenesis in Randomly Constructed Gene Nets, J. Theoret. Biol., (22), 437-467.

[12] S. Kauffman, (1993), The Origins of Order: Self-organization and Selection in Evolution, New York: Oxford Univ. Press.

[13] S. Kim, S. Imoto and S. Miyano, (2003), Dynamic Bayesian Network and Nonparametric Regression for Nonlinear Modeling of Gene Networks from time Series Gene Expression Data, Proc. 1st Computational Methods in Systems Biology, Lecture Note in Computer Science, (2602), 104-113.

[14] R. Pal, I. Ivanov, A. Datta, M. Bittner and E. Dougherty, (2005), Generating Boolean Networks with a Prescribed Attractor Structure, Bioinformatics, (21), 4021-4025.

[15] C. E. Shannon, (1948), A Mathematical Theory of Communication, Bell System Technical Journal, (27), 379-423.

[16] I. Shmulevich, E. Dougherty, S. Kim and W. Zhang, (2002), Probabilistic Boolean Networks: A Rule-based Uncertainty Model for Gene Regulatory Networks, Bioinformatics, (18), 261-274.

[17] I. Shmulevich and E. Dougherty, (2007), Genomic Signal Processing, Princeton University Press, U.S.

[18] S. Zhang, W. Ching, X. Chen and N. Tsing, (2010), On Construction of PBNs from a Prescribed Stationary Distribution. Information Sciences (180), 2560-2570.