

A COMPUTATIONAL MODEL FOR SIMULATING CONTINUOUS TIME BOOLEAN NETWORKS

Hakan Öktem, Ronald Pearson, Olli Yli-Harja,
Daniel Nicorici, Karen Egiazarian, Jaakko Astola
Tampere University of Technology, Institute of Signal Processing
P.O. Box 553, Tampere 33101, Finland
e-mails : oktem@cs.tut.fi, pearson@cs.tut.fi

ABSTRACT

Random Boolean networks are among the most popular model systems used for modelling the gene regulatory networks due to their answering of the many biological questions in a realistic way, insights into the overall behavior of large genetic networks and suitability for inference. However, discrete time Boolean networks have serious limitations in simulating non-repeating complex dynamic systems and incorporating the biochemical information on the reaction rates. In this work we introduce continuous-time Boolean network structure to simulate the genomic regulation as a binary complex dynamic system, discuss their potential of solving the problems of the conventional discrete time Boolean networks and develop a framework for simulating continuous time Boolean networks.

1 Introduction

Inference of gene regulation mechanisms became an emerging interest of research due to two important developments in genomics. One is the introduction of expression microarray technology, which allow following the genome wide transcriptional activity [1][8][11]. Another is the completion of many genome sequencing projects concentrating the interest on capturing the functions of genes in the whole context of cellular mechanisms [10]. Transcription of genes regulates the transcription factors and transcription factors regulate the transcription of the genes. This constitutes a dynamic regulatory network [2][5][6][12]. There have been various attempts to model gene regulatory networks. One of the model systems which received high interest was the *Random Boolean Network* model introduced by Kauffman [7]. Boolean Networks have several advantages with respect to other models. Many realistic biological questions may be answered within Boolean formalism rather than quantitative biochemical details [11]. This is the only model that has yielded insights into the overall behavior of large genetic networks [11][13]. The model is to some extent capable of demonstrating the functional differentiation of different

types of cells with the same genome [13] and it allow inferring the structure of genomic networks from experimental data [11][1]. However, this model has also important limitations on describing the genomic regulatory networks due to its discrete time structure. First of all discrete time Boolean networks converge to attractors of self repeating patterns [13]. However, yeast cell cycle studies demonstrate that the genomic regulation has a complex dynamic characteristic disproving a self repeating activity [8][9]. Second, the inferred network and underlying Boolean relations between the genes will depend on the sampling period, which may result with a different structure for the same network at different sampling rates. Third, the biochemical information on the reaction rates can not be incorporated into the model. Finally, the *Random Boolean Network* model provides a way of relating cell-specific information to patterns of state variables (gene expression profiles). A very similar concept was also discussed in neurophysiology, where short-term dynamic memory has been modeled using patterns of self re-exciting impulse trains [4]. However, finite state space divided into basins of attraction, converging to self repeating attractors is neither compatible with the cell specific dynamic information which a complex dynamic (chaotic) system can exhibit nor adequate to exhibit realistic cell-to-cell and temporal variations in genomic activity. Further, a conflict is likely to arise in Boolean networks between adaptivity and stability [13], both of which are required for cell viability.

We introduce *Continuous-Time Boolean Networks (CTBN)* to simulate the genomic activity as a binary complex dynamic system. A CTBN is a network of state variables whose states are binary and the relations between the state variables are given by Boolean relations with delays. This networks are also known as *Boolean Delay Equations (BDE)*. Dee and Ghil (1984) studied the behavior of the systems formed by BDE[3]. A *Binary Complex Dynamic System* is a system of binary variables which is non-repeating in Phase Space, the space whose dimensions are lags of the original variables. This behavior is possible in systems based on binary variables evolving in continuous

time, and this formulation is realistic if we view changes in gene expression as discrete events that can occur at any time. Furthermore, it is possible to incorporate some of the information used in concentration-based models into this formulation. That is, knowledge of how discrete changes in gene expression levels influence rates of change in protein concentrations and how gene expression is influenced by protein concentrations exceeding critical threshold values may be used to determine how long it takes a change in a binary state variable to influence other binary state variables.

2 Continuous-Time Boolean Networks

The discrete time Boolean network consisting of k state variables can be expressed by the following state transition equations,

$$s_{i,n+1} = f_i(s_{1,n}, s_{2,n}, \dots, s_{k,n})$$

where, $s_{i,n}$ is the binary state of i^{th} state variable at n^{th} time sample and f is a binary function of the previous values of the state variables.

In this scheme

$$S_n = [s_{1,n}, s_{2,n}, \dots, s_{k,n}]$$

there exist $m, m < 2^k$ such that $S_{n+m} = S_n$

$$S_{n+m} = S_n \Rightarrow S_{n+m+1} = S_{n+1}$$

due to the finite number of possible states. Hence a discrete time Boolean network will run into a self repeating state. Here S_n denote the state vector.

Now let us consider a continuous time Boolean network with delays:

$$s_{i,t} = f_i((s_{1,t-\tau_{1,i}}), (s_{2,t-\tau_{2,i}}), \dots, (s_{k,t-\tau_{k,i}})),$$

here $\tau_{j,i}$ is the delay which is the time necessary for a change in the state of variable j affect the variable i . Now we can form non-repeating patterns while the system oscillates but does not repeat itself exactly. In this case we need to keep track of the state variable in continuous time. Note that we can express each state variable by at least two quantities to trace the continuous time binary system by a discrete time computation method.

$$s_{i,n} = (s_{i,nt_s}, d_{i,nt_s}),$$

Where t_s is the sampling period, s_{i,nt_s} is the binary value of the state of the i^{th} variable at the n^{th} sample and d_{i,nt_s} is the continuous value indicating the time past since the last change in the state of the i^{th} variable at time nt_s . Since

the values are binary we can keep the track of the system in continuous time by keeping and recalculating the state change times of the variable. If the sampling period t_s is kept sufficiently small with respect to the reaction times the possibility of multiple state changes within the sampling period will be eliminated. If the delays in the system are close to each other only keeping the time of the last state change information is sufficient to keep sufficiently long history information. In case some variables may change very rapidly with respect to the delays in its output to affect the other variables more than one last state change time information should be kept to have sufficiently long memory. In practice keeping at least last two state change times will prevent possible errors caused by multiple state changes within the delay time.

The system may repeat a previous state but it does not need to cause the repetition of the full trajectory since the past continuous time after the last state changes do not need to be same. Complex dynamic systems have the potential of conveying two properties

1. Each of x different systems with the same structure may have unique trajectory based on initial states, without need of continuous external control variables, which may be associated with different varying genomic activity of different types of cells with the same genome.
2. Systems, which will never repeat their trajectories, can be built and any past effect on such a system may have a unique trace on the systems trajectory and still the system may continue oscillating around the same attractor. This property may be associated with functional differentiation in time and an adaptivity property.

2.1 Example

We would like to demonstrate the above mentioned two properties by a very simple example. Let us consider a continuous time binary network of 5 state variables (s_1, s_2, \dots, s_5) and 4 external inputs (e_1, e_2, e_3, e_4)

$$s_1(t) = (\overline{e_1}(t) \wedge \overline{s_2}(t - 1.1)) \vee (e_1(t - \tau_1) \wedge e_2(t))$$

$$s_2(t) = \overline{s_1}(t - \pi + 2)$$

$$s_3(t) = (\overline{e_3}(t) \wedge \overline{s_4}(t - 1.1)) \vee (e_3(t - \tau_1) \wedge e_4(t))$$

$$s_4(t) = \overline{s_3}(t - 1.11)$$

$$s_5(t) = s_2(t - 1.07) \wedge s_4(t - 1.07)$$

in this system external inputs can carry s_1 and s_3 to any initial state while s_2 and s_4 can be controlled by the past values of s_1 and s_3 respectively. Let us consider s_1, s_2, s_3, s_4 are switched off at time $t = -0.05$ and the influence of the external inputs are inactivated for time $t > 0$ by setting $e_1(t)$

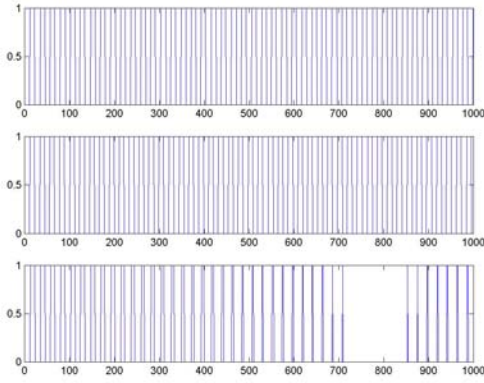


Fig. 1. The binary states of s_2 (up), s_4 (middle) and s_5 (down). $t_s = 0.1$.

and $e_3(t)$ to 0 for $t > 0$. When both s_1 and s_2 switched off at the same time t_0 , s_1 will be switched on at $t = t_0 + 1.1$ and s_2 will be switched on at $t = t_0 + \pi - 2$. At time $t = t_0 + \pi - 0.9$ both variables will be switched off again. Hence s_1 and s_2 will converge to an oscillation with a period $\pi - 0.9$. s_3 and s_4 will again converge to an oscillating state with a period 2.21 in the same way. Since π is a non-rational real quantity the time between the onsets of s_2 and s_4 will never repeat. Therefore, the trajectory of s_5 will be non-repeating and still its periodicity would be kept within the same limits. The states of s_2 , s_4 and s_5 are shown in figure 1 and the last state change times of s_5 is shown in figure 2. Again if one of the oscillating pairs is shifted in time the trajectory of s_5 would be shifted but it would still continue a similar behavior. However, if the initial state of s_1 is set to 1 and s_2 is set to 0 and kept the same for a period longer than $\pi - 2$, s_1 will remain on and s_2 will remain off. Hence s_5 will never be on.

An important issue to consider here is the possible amount of initial conditions, which the systems can encounter. In **CTBN** not only the states of the variables but the time between the state changes as well are distinctive in determination of the future behavior of the system. Since the time is allowed to be a continuous variable, a **CTBN** can handle dynamic information in continuous scale while the possible number of initial conditions is limited with 2^k in **DTBN**.

2.2 Algorithm

For preventing multiple state changes within the sampling period we assume time limits for a gene can its state. Considering reactions turn the genes on or off (binding of promoters or inhibitors) also need some time this assumption

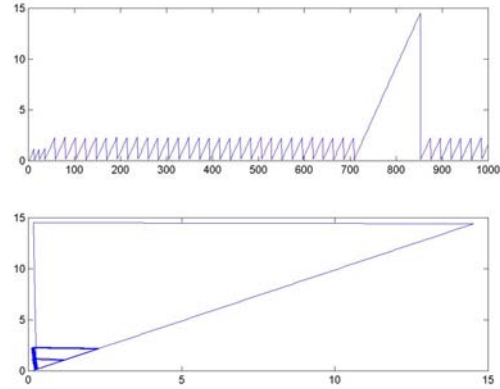


Fig. 2. The last state change times of s_5 in time (up) and its phase plot (down). $t_s = 1$.

will be realistic depending on the selected sampling period for simulation.

The basic simulation algorithm consists of the following steps:

1. For each time sample n and for each gene i , read the Boolean function to compute the variable $s_{i,n}$.
2. Compare the last state change time information and corresponding delays of each variable affecting the state of i . Use the states of variables for the ones no change has occurred in their state for a period longer than the corresponding delay. Otherwise, use the inverse of the states which has changed sooner than the delay.
3. Incorporate external effects if any external control input has applied.
4. If no change occurs in the state of the variable s increase the corresponding last state change time information by the sampling period, otherwise calculate the new last state change time information by tracing the variables affecting its state and finding the time of the first change switching i on or off.

If more than one last state change time information is used, the algorithm can be extended by repeating the step 2 for each of the older state change times and finding the older times at step 4. Since the state change times except the last are already computed, only shifting the times increased by the sampling period to an older state is sufficient.

3 Discussions and Conclusions

Here we will discuss two questions: whether we can identify these models from experimental data, and whether we can incorporate biochemical information into these models.

One of the important advantages provided by Discrete Time Boolean Networks (**DTBN**) is the possibility of identifying them from experimental data[1][11][13]. When we introduce continuous time delays continuous time signal is necessary (but may not be sufficient) for inferring on the basis of experimental data. However, if the delays can be found or estimated the inference has the same complexity with the **DTBN**. Furthermore, it eliminates the dependency of the inferred structure to the sampling rate. Considering **DTBN** is a special version of **CTBN** at all delays are equal to the sampling period, the inaccuracy in the estimated delays has the most risk of providing same results.

Another popular model system is concentrated on simulating particular regulatory pathways based on incorporating the quantitative biochemical details and providing combined models of gene states and protein concentrations[2][6][10][12]. Some of those models are considering the states of the genes determined by the protein concentrations over or under a threshold[2][12]. Even considering single reactions within a single cell as a continuous event is discussed[13], we can presume that the delays within single reactions in a single cell is very highly related with the continuous variable reaction kinetics observed in a culture of several cells. Since the available information is levels of mRNA, the verifiability of this type of information is also important. Hence, we can determine some of the delays by using the available reaction kinetical information. Considering the protein concentrations depend on how long the genes are on [13] we assume that the concentrations may be computed from the timing of the gene states.

In this work we introduced a novel model structure for representing the genomic regulation as a *Continuous-Time Boolean Network* and proposed a framework for simulating a **CTBN** model. The system has the potential of simulating a wide range of the features of genomic regulation including the complex dynamic characteristics.

Acknowledgement 1 *We thank Professor Ilya Schmulevich from MD Anderson Cancer Research Center Austin Texas USA for the valuable discussions about the subject.*

4 References

[1] T. Akutsu, S. Miyano, "Identification of Genetic Networks from a Small Number of Gene Expression Patterns under the Boolean Network Model" Proc. Pacific Symposium on Biocomputing'99 (PSB'99), 17-28, 1999.

[2] K.C. Chen, et.al. "Kinetic Analysis of a Molecular Model of the Budding Yeast Cell Cycle" Molecular Biology of the Cell, Vol 11, 369-391, January 2000.

[3] Dee D. and Ghil M. Boolean difference equations, i: Formulation and Dynamic behavior. *SIAM J. Appl. Math.*, 44:111-126, 1984.

[4] J.C. Eccles, "The understanding of the Brain", The Johns Hopkins Press, 1972

[5] A.J. Hartemink, D.K. Gifford, R.A. Young, "Combining Location and Expression Data for Principled Discovery of Genetic Regulatory Network Models" Pacific Symposium on Biocomputing 2002, Kauai, January 2002

[6] N.S. Holter, A. Maritan, M. Cieplak, N.V. Fedoroff, J.R. Banavar, "Dynamic Modeling of Gene Expression Data" Proc. National Academy of Sciences, 98(4):1693-1698, 2001

[7] S.A. Kaufmann, "The Origins of Order: Self Organization and Selection in Evolution", Oxford University Press, New York, 1993.

[8] R.R. Klevecz and D.B. Murray, "A Deep Basin of Attraction in the Yeast Transcriptome Uncovered by Wavelet Decomposition" Proc. Workshop on Nonlinear Signal and Image Processing, 2001.

[9] R.R. Klevecz, S.A. Kauffman, R.M. Shymko, "Cellular clocks and oscillators.", International Review of Cytology, 86:97-128. 1984.

[10] H. Matsuno, et.al "Genomic Object Net: Basic Architecture for Representing and Simulating Biopathways" 9th International Conference on Intelligent Systems for Molecular Biology, 2001.

[11] I. Shmulevich, O. Yli-Harja, J. Astola, "Inference of Genetic Regulatory Networks Under the Best-fit Extension Paradigm" Proc. Workshop on Nonlinear Signal and Image Processing, 2001.

[12] I. Simon, et.al. "Serial Regulation of Transcriptional Regulators in the Yeast Cell Cycle" Cell 2001 106: 697-708.

[13] A. Wuensche, "Genomic Regulation Modeled as a Network with Basins of Attraction", Proc. 1998 Pacific Symposium on Biocomputing, 89-102