



Scalable approach for effective control of gene regulatory networks

Mehmet Tan^{a,b}, Reda Alhajj^{b,c,*}, Faruk Polat^a

^a Department of Computer Engineering, Middle East Technical University, Ankara, Turkey

^b BIDEALS Group, Department of Computer Science, University of Calgary, 2500 University Drive NW, Calgary, Alberta Canada

^c Department of Computer Science, Global University, Beirut, Lebanon

ARTICLE INFO

Article history:

Received 21 April 2008

Received in revised form 25 September 2009

Accepted 3 October 2009

Gene regulatory networks
Control policy
Scalability technique
Feature reduction
Probabilistic Boolean networks
Markov decision problems

ABSTRACT

Objective: Interactions between genes are realized as gene regulatory networks (GRNs). The control of such networks is essential for investigating issues like different diseases. Control is the process of studying the states and behavior of a given system under different conditions. The system considered in this study is a gene regulatory network (GRN), and one of the most important aspects in the control of GRNs is scalability. Consequently, the objective of this study is to develop a scalable technique that facilitates the control of GRNs.

Method: As the approach described in this paper concentrates on the control of GRNs, we argue that it is possible to improve scalability by reducing the number of genes to be considered by the control policy. Consequently, we propose a novel method that considers gene relevancy to estimate genes that are less important for control. This way, it is possible to get a reduced model after identifying genes that can be ignored in model-building. The latter genes are located based on a threshold value which is expected to be provided by a domain expert. Some guidelines are listed to help the domain expert in setting appropriate threshold value.

Results: We run experiments using both synthetic and real data, including metastatic melanoma and budding yeast (*Saccharomyces cerevisiae*). The reported test results identified genes that could be eliminated from each of the investigated GRNs. For instance, test results on budding yeast identified the two genes SWI5 and MCM1 as candidates to be eliminated. This considerably reduces the computation cost and hence demonstrate the applicability and effectiveness of the proposed approach.

Conclusion: Employing the proposed reduction strategy results in close to optimal solutions to the control of GRNs, which are otherwise intractable due to the huge state space implied by the large number of genes.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Protein synthesis is a key process for living organisms. All proteins are encoded by messenger RNA (mRNA), which is extracted from a gene in DNA; proteins are produced in a process that involves two stages, namely transcription and translation. In transcription, a sequence of the gene is used to produce mRNA, which is then used to create a protein during translation. For a gene to be transcribed into mRNA, it is often necessary for a specific protein called *transcription factor* to bind to the DNA in a specific location. A transcription factor can have a positive or negative regulatory effect on the binding site. So, the transcription level (or

the expression level) of the gene can change based on the binding of the transcription factor. Since the transcription factor is also a protein, which is decoded from a gene, it is possible to describe and discuss a set of interactions among genes; these interactions constitute a GRN.

As described in the literature, there are various methods to represent and model a GRN [1]. These include (dynamic) Bayesian networks, (probabilistic) Boolean networks (BNs), neural networks, petri-net models and differential equation-based models. Modeling may provide an opportunity to estimate the future state of a cell based on the current state and the conditions affecting the cell. To justify the need for control, consider a cell which is estimated to be in an undesirable state (e.g., cancerous state) in the near future; this brings the necessity to intervene the current state of the network in order to avoid reaching undesirable state(s). But, it is important to intervene as efficiently and effectively as possible because of the urgency of the situation and the cost of the intervention. This motivates for the need to control GRNs, the problem may be stated as follows: *find an efficient policy to interact*

* Corresponding author at: BIDEALS Group, Department of Computer Science, University of Calgary, Calgary, Alberta Canada. Tel.: +1 403 220 9453; fax: +1 403 284 4707.

E-mail addresses: mtan@ceng.metu.edu.tr (M. Tan), alhajj@ucalgary.ca (R. Alhajj), polat@ceng.metu.edu.tr (F. Polat).

(by interventions) with the network in order to change the behavior in a way that satisfies some prespecified objective(s). On the other hand, the size of the state space is the most crucial issue in GRN control; this consideration is common to all control problems. Here it is also worth mentioning that the term control in the context of GRN is slightly different from control theory because of the limitedness of possible intervention means for GRNs.

For a discrete GRN (where the expression levels of genes are discretized), the size of the state space is proportional to the number of genes and the number of levels of discretization for each gene. Even if the expression levels of the genes are discretized to binary levels, the size of the state space is 2^N for an N -gene network; this makes the problem hard to cope with even for small values of N . So, to find an efficient policy for the GRN control problem, appropriate methods must be introduced to reduce the state space to a reasonable size, whenever possible and desired.

The relevancy of a given gene in terms of control depends on the objective to be satisfied because genes in a GRN have varying effects on each other; this means that a gene might have minimal or negligible effect on the solution of a GRN control problem. In this paper, we utilize the relevancy measure to propose a kind of feature reduction method capable of identifying genes which are less relevant for control. Such genes are candidates to be eliminated in building a model so that an approximate control policy can be reached faster. The feature reduction process may identify more than one gene as candidates to be eliminated. But, even when one gene is eliminated, the state space reduces significantly. Obviously, this positively reflects on the scalability of the GRN control problem to be investigated.

Generally, the control of GRNs has been studied on Markovian models, e.g., [2–10]. For instance, Shmulevich et al. [10] considered control in a Markovian model by exploiting the Markov chain theory [11]. It has been shown how to select the gene to intervene in order to minimize the time required to reach some set of desirable states, given the current state. Structural intervention is also considered for reaching desired states [9]. On the other hand, Datta et al. [5] formulated the interventions in terms of altering transition probabilities by using some external control variables. They used dynamic programming to formulate and solve a finite horizon controlled Markov chain, where a horizon is the duration of applying external actions and the Markov chain is defined similar to a Markov decision process [12]. Optimal infinite-horizon control extension of this work is described in [8].

Almost all the above mentioned studies use probabilistic Boolean networks (PBNs) [13] as the Markovian model. A slightly different model in the context of control is investigated in [7], where the switching between the BNs forming a probabilistic Boolean network (PBN) is not performed in every step, but probabilistically. Attractors in a PBN are the states to be reached after a finite number of steps and this stable situation will not change in the absence of perturbations. Control with the lack of the knowledge of switching probabilities between BNs that have common attractors has been studied by Choudhary et al. [4].

A previous study in our group [2,3] is based on the following argument: *if the control process is considered as a treatment then observing the patient after the treatment can also be taken into account while solving the control problem*. As a result, the solution described in [2] was developed; it considers a monitoring horizon after the control horizon. The solution is given for various settings depending on the control and monitoring horizons being finite or infinite. The problem was also formulated as a multi-objective problem where the objectives are state cost and state-action cost defined by domain experts [3].

To sum up, the above mentioned works focus on solving the control problem in GRNs for different settings using dynamic

programming. A gene is assumed to be relevant if it is chosen for modeling, i.e., it exists in the same GRN with others. But, we observed that the relevancy also depends on the objective(s); and consequently we argue that the component of the GRN we should focus on may change according to the given set of objectives. Based on this argument, we propose a feature reduction method that successfully maintains scalability in the control of GRNs [14,15]. By feature reduction, we provide the choice to reduce the number of genes to be considered in the control process and hence maintain scalability. Neglecting scalability turns control into an unmanageable process, though control is essential to study and understand the behavior of any given system. To the best of our knowledge, this is a major contribution as the first attempt of applying feature reduction in the context of GRNs; our initial results have encouraged us to expand the work as described in this paper. The results reported in this paper demonstrate the applicability and effectiveness of the proposed approach. Although GRN control studies are not yet directly applicable to clinical practice, the promising results demonstrate the potential to be used in real applications. We reported test results using both synthetic and real gene expression data.

The rest of this paper is organized as follows. Section 2 includes the necessary background information. Section 3 covers the details of the proposed reduction based approach. Section 4 reports experimental results on synthetic and real gene expression data. Section 5 is conclusions and future research directions.

2. Background

In this section, we cover the background necessary for the scope of the work described in this paper. In particular, we present an overview of the Markov decision problems (MDPs) and discuss the control problem in the context of GRNs.

2.1. Markov decision problems

A Markov decision process is formally defined as a quadruple (S, A, T, R) , where S is the set of states, A is the set of actions, T is transition probabilities such that $T(s, a, s')$ denotes the probability of the next state being s' when action a is applied in state s , and R is the real-valued reward function defined on the set $S \times A$. R is sometimes defined on the set $S \times A \times S$ as $R(s, a, s')$, where the next state also effects the reward function. This definition can be easily transformed into the former as, $R(s, a) = \sum_{s'} T(s, a, s') R(s, a, s')$. A Markov decision process with a performance criterion is called a Markov decision problem (MDP). Solution to a MDP is a mapping $\pi : S \rightarrow A$, which is called *policy*. The aim here is to find a policy that maximizes the total expected future reward, which is defined differently based on the performance criterion. Unless otherwise stated, we focus in this paper on total discounted infinite horizon future whose definition can be specified as: $\sum_t \beta^t R_t(s, a)$, where $R_t(s, a)$ is the immediate reward on performing action a in state s at time step t , and the discount factor β is in $(0, 1)$.

The value function for state s , denoted $V(s)$, gives the value of desirability of state s for the current control process. Every policy π defines a value function V^π on state space S . The value of a policy π in state s is the total reward of choosing an action in state s according to π , and following π thereafter. The Bellman equation [16] defines the relationship between $V^\pi(s)$ and values of other states:

$$V^\pi(s) = R(s, \pi(s)) + \beta \sum_{s'} T(s, \pi(s), s') V^\pi(s') \quad (1)$$

There exists a unique optimal value function [12], denoted V^* , which is the value function of the optimal policy π^* . The optimal

Download English Version:

<https://daneshyari.com/en/article/377904>

Download Persian Version:

<https://daneshyari.com/article/377904>

[Daneshyari.com](https://daneshyari.com)