

Applying data mining to learn system dynamics in a biological model

Bingchiang Jeng^{a,*}, Jian-xun Chen^a, Ting-peng Liang^b

^a Department of Information Management, National Sun Yat-sen University, 70 Lien-hai Road, Kaohsiung City 804, Taiwan, ROC

^b Department of Information Management, National Sun Yat-sen University, Taiwan

Abstract

Data mining consists of a set of powerful methods that have been successfully applied to many different application domains, including business, engineering, and bioinformatics. In this paper, we propose an innovative approach that uses genetic algorithms to mine a set of temporal behavior data output by a biological system in order to determine the kinetic parameters of the system. Analyzing the behavior of a biological network is a complicated task. In our approach, the machine learning method is integrated with the framework of system dynamics so that its findings are expressed in a form of system dynamics model. An application of the method to the cell division cycle model has shown that the method can discover approximate parametric values of the system and reproduce the input behavior.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Systems biology; System dynamics; Neural network; Genetic algorithm

1. Introduction

Traditionally, researchers adopt the reductionism to study biological phenomena, i.e. analyzing a system by breaking it into constituents repeatedly until they can be observed directly (Gallagher & Appenzeller, 1999). In order to find the function and role of a component, the researcher has to repeatedly conduct experiments with different system parameters or components. Although this approach works fine in most situations, it often encounters difficulties when we intend to examine the interaction effects within a system or when the system is complicated. It is also well-known that the net behavior of a biological system is usually not the sum of its components' behavior (Csete & Doyle, 2002) because of the existence of the so-called 'emergent property' (Bhalla & Lyengar, 1999; Gardner & Collins, 2000; Yi, Huang, Simon, & Doyle, 2000).

Recently, a system view of biology called *systems biology* has been proposed (Chong & Ray, 2002; Davidson, Rast, Oliveri, Ransick, Caletani and Yuh, 2000; Kitano, 2002a), which aims to the development of a system-level understanding of biological systems (Kitano, 2002a). In other words, one wants to understand not only the molecules but also

the cause–effect relationships linking the behavior of molecules as well as the characteristics and functions of a system. Although artificial intelligence has increasingly been used in analyzing biological data for years, this is certainly a more difficult case and needs innovative methods.

We propose an approach that integrates system dynamics and data mining methods to induce the dynamic behavior of a biological system in this paper. System dynamics is a discipline that studies the dynamic behavior of social systems (Forrester, 1961). In particular, it has an advantage in modeling the information-feedback characteristics to see how system structure, amplification (in policies), and time delays (in decisions and actions) may interact to influence the behavior of an organization. Since a social system is a combination of a number of simple entities (or agents) that operate in an environment to generate complex behavior patterns as a collective, it may be suitable for analyzing the information-feedback loops and complicated interactions within a biological system (Becskei & Serrano, 2000; Gardner & Collins, 2000).

A challenge for applying system dynamics to the analysis of biological data is that the base model for analysis is often constructed by human experts who have expertise in the application domain and are able to draw a flow diagram by observing the operation of target system to represent the causal relationships among system entities (variables) (Coyle, 1977; Lyneis & Pugh, 1996; Starr, 1980). This, however, is not the case in biological analysis because in most cases the biological systems under study act like black-boxes and only their input and output behavior can be observed over time. Thus, direct

* Corresponding author. Tel.: +886 7 525 2000x4719; fax: +886 7 525 4799.

E-mail addresses: jeng@mail.nsysu.edu.tw (B. Jeng), jamesc02@ms9.hinet.net (J.-x. Chen), tpliand@mail.nsysu.edu.tw (T.-p. Liang).

application of system dynamics to the construction of biological models is very difficult, if not impossible. We need a mechanism to bridge the gap.

A possible way to deal with the problem is to use data mining techniques to analyze the observed behavior data to discover the hidden relationships and/or rules behind the system dynamics. In order to do this, a data mining method needs to be augmented; it has to have a conceptual framework beforehand so that the findings from data will be express in the form of a system model. In this paper, we will use a combination of genetic algorithms and artificial neural networks to implement the idea. The artificial network is designed to emulate a system dynamics model and then encode into a genetic form for learning. The proposed approach is applied to experiment on the synthetic cell division cycle model (CDC6, hereafter) created by Tyson (1991). The behavior data generated by CDC6 model is given as an input to the developed method to learn the model's kinetic parameters. The results are then compared with the original data to evaluate the effectiveness of the approach.

The remainder of the paper is organized as follows. Section 2 is a brief review of related literature. Section 3 describes the proposed approach for mining behavior relationships from a set of observed biological data. Section 4 illustrates the result when the approach is applied to the CDC6 model. Section 5 concludes the paper.

2. Review of literature

2.1. Systems biology

One thing that realizes systems biology is the high-throughput measurement devices for DNA, RNA, and proteins. The ascendancy of these high-throughput devices in the past decade has permanently changed the biological landscape of genomics, proteomics, and metabolomics studies (Henry & Washington, 2003). Rising as a new star under this background, systems biology aims at a system-level understanding of biological systems (Kitano, 2002a). Unlike molecular biology, which focuses on the study of molecules such as nucleotide acids or protein sequences, systems biology focuses on dynamics of systems, which cannot be described merely by enumerating the molecular components of the system (Henry & Washington, 2003). Another misleading concept in molecular biology is to believe that only system structure, e.g. network topologies, is important without paying attention to the diversities and functionalities of system components. Both the structure and the components are indispensable in forming the symbiotic state of a system as a whole (Henry & Washington, 2003).

Research of systems biology focuses on four key properties of a biological system: (1) system structure, (2) system dynamics, (3) control method, and (4) design method (Kitano, 2002a). System structures include networks of gene (or protein) interactions, biochemical pathways, and the mechanisms to modulate the physical properties of intra- and multi-cellular structures.

System dynamics concerns a system's dynamics behavior over time under various conditions. There are currently several

methods to analyze the dynamics of a system from different perspectives: metabolic analysis, sensitivity analysis, phase portrait analysis, bifurcation analysis, and analysis by identifying essential mechanisms of a specific behavior (Kitano, 2002a). The typical one is bifurcation analysis, which traces the time-varying state changes of the system with a multi-dimensional plot(s) where each dimension represents a particular biochemical concentration involved in the interaction.

Control methods are related to ways to control the states of a biological system, e.g. the methods to transform cells from malfunctioning into healthy ones. It is an application that makes the knowledge obtained from system structure and system dynamics. Design methods move even further. It intends to establish technologies to design biological systems in a way we wish. An example is the attempt to actually design and grow organs from the patient's own tissue (Kitano, 2002a).

Research of systems biology is emerging as an area of potential (Chong & Ray, 2002; Kitano, 2004; Nobel, 2002). For example, Kitano (2004) recently puts forth a new cancer treatment proposal from the viewpoint of systems biology. Cancer disease has a very robust system, which causes many medical treatments fail to control the growth of cancer cells. Thus, it is hard to defeat a cancer by simply alternating some constituent of the system, but it may be possible to put cancer cells into the status of dormancy or apoptosis, if one can change the kinetic parameters (rates) of the system from a system's viewpoint. Recognizing the importance and potential of systems biology, a special issue on this subject has been published in the March 2002 issue of *Science*.

2.2. System dynamics models

As mentioned above, system dynamics concerns a system's dynamics behavior over time under various conditions. Its emergence dates back to the publication *Industrial Dynamics* by Jay W. Forrester early in 1961 (Forrester, 1961). In his book, Forrester defines system dynamics as "the study of the information-feedback characteristics of industrial activity to show how organizational structure, amplification (in policies), and time delays (in decisions and actions) interact to influence the success of the enterprise."

A system dynamics model is one that models the dynamics of such a system. There are some different ways to do this and the most typical one is Forrester's flow diagram (Forrester, 1961). As an example, the elementary inventory control system (see Fig. 1) described in Forrester's book is redrawn here to illustrate the notation and modeling concepts.

For simplicity, assume that the order rate (*OR*) in the model can be either positive or negative, i.e. goods can either be ordered into the inventory or returned back to a supplier. The goal is to maintain the inventory at a fixed level defined by the desired inventory (*DI*) for duration of time. In order to bring the actual inventory toward the desired level, one has to increase the order rate when the inventory falls far below the desired value, and, as it approaches the desired inventory, the order rate should be gradually reduced. If, on the other hand, the inventory becomes

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات