BioSystems 153–154 (2017) 6–25

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

A system of recurrent neural networks for modularising, parameterising and dynamic analysis of cell signalling networks

S. Samarasinghe*, H. Ling

Integrated Systems Modelling Group, Lincoln University, New Zealand

ARTICLE INFO

Article history:
Received 7 July 2016
Received in revised form 1 December 2016
Accepted 23 January 2017
Available online 4 February 2017

Keywords:
Recurrent neural networks
Cell signalling networks
Model decomposition
Modularisation
Cell cycle
G1/S transition

ABSTRACT

In this paper, we show how to extend our previously proposed novel continuous time Recurrent Neural Networks (RNN) approach that retains the advantage of continuous dynamics offered by Ordinary Differential Equations (ODE) while enabling parameter estimation through adaptation, to larger signalling networks using a modular approach. Specifically, the signalling network is decomposed into several sub-models based on important temporal events in the network. Each sub-model is represented by the proposed RNN and trained using data generated from the corresponding ODE model. Trained sub-models are assembled into a whole system RNN which is then subjected to systems dynamics and sensitivity analyses. The concept is illustrated by application to G1/S transition in cell cycle using Iwamoto et al. (2008) ODE model. We decomposed the G1/S network into 3 sub-models: (i) E2F transcription factor release; (ii) E2F and CycE positive feedback loop for elevating cyclin levels; and (iii) E2F and CycA negative feedback to degrade E2F. The trained sub-models accurately represented system dynamics and parameters were in good agreement with the ODE model. The whole system RNN however revealed couple of parameters contributing to compounding errors due to feedback and required refinement to sub-model 2. These related to the reversible reaction between CycE/CDK2 and p27, its inhibitor. The revised whole system RNN model very accurately matched dynamics of the ODE system. Local sensitivity analysis of the whole system model further revealed the most dominant influence of the above two parameters in perturbing G1/S transition, giving support to a recent hypothesis that the release of inhibitor p27 from Cyc/CDK complex triggers cell cycle stage transition. To make the model useful in a practical setting, we modified each RNN sub-model with a time relay switch to facilitate larger interval input data (≈20 min) (original model used data for 30 s or less) and retrained them that produced parameters and protein concentrations similar to the original RNN system. Results thus demonstrated the reliability of the proposed RNN method for modelling relatively large networks by modularisation for practical settings. Advantages of the method are its ability to represent accurate continuous system dynamics and ease of: parameter estimation through training with data from a practical setting, model analysis (40% faster than ODE), fine tuning parameters when more data are available, sub-model extension when new elements and/or interactions come to light and model expansion with addition of sub-models.

1. Introduction

The field of systems biology is concerned with understanding the principles of organisation and functionality of living systems for the purpose of advancing our knowledge of biology, health and diseases and developing effective cures. Biological systems can be characterised as hierarchically organised, interconnected and complex web of entities interacting in space and time for performing myriad functions to sustain life. A current emphasis in systems biology is the study of regulatory pathways- network of genes, proteins and other molecules linked by biochemical reactions. A key focus is to understand phenomena including homeostasis and instability that emerge from the way that constituent elements interact. The system, not the building blocks, is the subject matter. A bottleneck for analysing systems behaviour of regulatory pathways (networks) is that typically they are extensively cross-linked networks of multiple sub-pathways that are highly regulated through feedback loops. This leads to pathways without an easily discernible boundary where overlapping sub-pathways continuously share information in the temporal unfoldment of the process. Any aspect, for example cell cycle system or stages of cell cycle, can be considered as a functional subsystem embedded in a larger sys-

* Corresponding author.
E-mail address: Sandhya.samarasinghe@lincoln.ac.nz (S. Samarasinghe).

http://dx.doi.org/10.1016/j.biosystems.2017.01.001
0303-2647/

tem of networks as an integral and interlinked part of the whole. Full details of many such subsystems are yet to be characterised but they are intricate and complex at subsystem or any level. As more and more omics data and knowledge of pathways are gathered at a fast pace, effective approaches to model and investigate signalling networks as complex systems become paramount.

To address this challenge, computational and mathematical approaches have been combined with biological knowledge to gain insight into how living systems function (Aderem, 2005; Ge et al., 2003; Heath and Kavraki, 2009; Hübner et al., 2011; Hyman and Whither, 2011; Ideker et al., 2001; Kitano, 2002; Kohl et al., 2010; Molina et al., 2010; Oltvai and Barabasi, 2002). These approaches fall broadly into two types: continuous (such as ODEs) and discrete (such as Boolean) models. Additionally, hybrid, stochastic, Bayesian, network induction and soft systems methods have been developed to improve various aspects of these models (Ling et al., 2013). An overview, advantages and disadvantages of these approaches were presented in our previous work (Ling et al., 2013). These models have been used in modelling (Ling et al., 2013): *gene regulatory networks* (de Jong, 2002; Hecker et al., 2009; Karlebach and Shamir, 2008; Polynikis et al., 2009; Qi et al., 2010), pr *otein signalling networks* (Albert et al., 2009; Gilbert et al., 2006; Hardy et al., 2011; Hughey et al., 2010; Sachs et al., 2002, 2005; Woolf et al., 2005; Tenazinha and Vinga, 2011) and *metabolic networks* (Berthoumieux et al., 2011; Papin et al., 2003, 2004; Pozo et al., 2010; Rizk and Liao, 2009). However, gaining insights into a particular biological process may require integration of several of these network types; for example, cell cycle is a signalling network involving integrated action of gene regulation and protein–protein signalling to produce two identical cells.

The major problem with signalling networks is representing a large number of interactions in a single model for analysing their temporal dynamics accurately. Boolean approaches based on logic based operations can provide a qualitative view of a large system due its simplicity of approach. Hybrid and various extensions of Boolean have been proposed for characterising continuous dynamics of signalling networks but these still need refinement for capturing complex behaviour (Ling et al., 2013; Singhania et al., 2011). On the other hand, ODEs can provide a comprehensive quantitative view of temporal dynamic of the system in fine detail but simulating a large network with ODEs is complex due to the large number of ODEs with many unknown parameters (Ashyraliyev et al., 2009; Lillacci and Khammash, 2010; Sun et al., 2008; Xie et al., 2010). The task of modelling large signalling networks becomes more manageable and efficient if the whole network can be divided into meaningful modules that can be modeled individually and then assembled to represent the whole system. At the same time, if an efficient approach to parameter estimation is applied at the modular level, the parameter estimation of the whole system becomes easier. In consideration of the strengths and limitations of the above methods, seamlessly modularising a system and effectively parameterising it while preserving the attractive attributes of ODE representation can be valuable for large signalling networks.

In our previous work (Ling et al., 2013), we proposed a novel recurrent neural networks (RNN) concept based on ODE description for representing and parameterising signalling networks. We presented its methodological development and proof of concept by demonstrating its efficiency in parameterisation and analysis of temporal dynamics and robustness through a successful application to a small 3-element system- p53-Mdm2 oscillatory system in cell cycle. In particular, we demonstrated that the model and its outcomes are biologically representative. In essence, we transformed an ODE system into an RNN framework that enabled parameter estimation and simplified subsequent model analysis.

In this article, we propose to extend this approach to a larger system for representing it as a system of subsystems and demonstrate its effectiveness for seamlessly modularising and parameterising larger networks and gaining novel insights into the biological system concerned. The idea here is that modularisation simplifies the sub-model structure and therefore they train more quickly as there are fewer parameters to optimise and smaller number of equations to process within each sub-model. In typical modularisation, a task is divided into independent sub-tasks that are coordinated by a higher level controller to achieve the task. However, for an interacting self-organising system without *a priori* knowledge about how it can be decomposed, a clear division into strongly autonomous modules is not possible and any subsystem can overlap with other subsystems in time and space. Therefore, many possibilities exist for modularising signalling networks and our focus is not on proposing the best approach to modularisation but to develop an efficient approach to represent a chosen modularisation. The rationale for modularisation used here to demonstrate the extension of the RNN model is based on the key events at various stages of the temporal progression of the system where modules interact dynamically in these events. Benefits of the proposed RNN approach is the ease of sub-model development and parameterisation and integration of sub-models into one model to produce accurate system dynamics more quickly than the ODE system. Furthermore, amenability of parameters to adaptation allows the scope for fine tuning them as more accurate data become available. Also, it could allow an opening for exploration of potential evolution of long-term learning and memory of these networks as there is increasing perception that somatic cells also learn as neurons do in our brain (Levin, 2014).

The basic idea is that the whole system is divided into a number of modules that represent specific important events in the chosen pathway and are represented by modular recurrent neural networks. However, unlike standard recurrent networks, the networks developed in this study (and our previous study (Ling et al., 2013)) are exact representations of the interactions in the biological system of interest. Therefore, each of the weights in the modular networks corresponds to a kinetic parameter in the corresponding system of ODEs and the assembled system of recurrent neural networks, simulated as a dynamical system, can provide vital biological insights into the behaviour of the actual biological system. In essence, we seek to transform a larger ODE system into a modular RNN system that enables modular level parameter estimation and simplifies the analysis of large signalling networks.

1.1. Recurrent neural network (RNN) for signalling networks

In our previous article (Ling et al., 2013), we highlighted that with the rapid rate of generation of omics data that are becoming more and more accurate and comprehensive, there is a high potential for Artificial Neural Networks (ANN) to contribute to systems biology. In that paper, we also gave a detailed overview of ANN in general and a type of recurrent networks (RNN) that shows much potential for representing, parameterising and analysing system dynamics of signalling networks as well as for modularising large networks. We demonstrated its application to a small 3-protein system. Here, we provide a very brief overview of ANN and the RNN developed in our previous work (Ling et al., 2013) that is adequate for the current purpose of highlighting how the method is extended to a larger network.

Two key attractions of ANNs are that they are mathematical models that can simulate nonlinear systems to an arbitrary degree of precision and they can solve a variety of complex problems that cannot be solved analytically by most mathematical models (Samarasinghe, 2006). This is achieved by a network of computational neurons (Fig. 1a) that collectively process information (inputs, x_i) to reach a desired output based on an iterative update of model parameters (weights, w_{ij}) during network training with