Identification of Functional Connections in Biological Neural Networks Using Dynamic Bayesian Networks

Chaoyi Dong*, Hong Yue**

** School of Electric Power, Inner Mongolia University of Technology, Huhhot, 010080, China*
(e-mail: dongchaoyi@imut.edu.cn)
*** Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow,G1 1XW, UK*
(e-mail: hong.yue@strath.ac.uk)

Abstract: Investigation of the underlying structural characteristics and network properties of biological networks is crucial to understanding the system-level regulatory mechanism of network behaviors. A Dynamic Bayesian Network (DBN) identification method is developed based on the Minimum Description Length (MDL) to identify and locate functional connections among Pulsed Neural Networks (PNN), which are typical in synthetic biological networks. A score of MDL is evaluated for each candidate network structure which includes two factors: i) the complexity of the network; and ii) the likelihood of the network structure based on network dynamic response data. These two factors are combined together to determine the network structure. The DBN is then used to analyze the time-series data from the PNNs, thereby discerning causal connections which collectively show the network structures. Numerical studies on PNN with different number of nodes illustrate the effectiveness of the proposed strategy in network structure identification.

Keywords: Pulsed Neural Networks (PNN); Dynamic Bayesian Network (DBN); Minimum Description Length (MDL); causal connection; synthetic biological networks.

1. INTRODUCTION

40 pt
0.556 in
14.1 mm

The molecular regulatory mechanisms in neuronal systems have been studied intensively, and the basic processes of information handling by neurons have been well understood (Bliss and Collingridge, 1993; Schuldiner et al., 1995; Trimble et al., 1991). However, little is known about the organization and function of complex biological neural networks (NNs) due to inadequate non-invasive measurement means or lack of effectiveness in data analysis methods. Experimental methods can simultaneously monitor electrical activities of a large number of neurons in real time thus providing multi-channel neural spike training data with sufficient temporal and spatial resolution. This was achieved by multiple single neuron recordings (Kruger, 1983), voltage sensitive dyes (Parsons et al., 1991), or multi-electrode array (Erickson, 2008; Kang et al., 2009; Spira and Hai, 2013).

Experimental data with several quantities simultaneously recorded produce multivariable datasets. Such multichannel recordings are intended to deliver more causal information about the investigated biological networks. Therefore, data-based identification of connective structures and further examination of the underlying regulation mechanisms are central themes in neural science and computational biology research. A number of modelling methods on describing biological network structures have been reported, such as Boolean network (Shmulevich et al., 2002), differential equation modelling (Kim et al., 2007), the likelihood ratio test method (Caines and Chan, 1975), and Granger causality (Blinowska et al., 2004; Doerfler et al., 2013). These methods have been applied in modelling of the causal dependency among genetic regulatory networks, protein-protein networks, and metabolite networks, etc. Nevertheless, there are fundamental limitations in applying these methods to biological NNs since most methods require detailed a priori knowledge on the system order and structure which could be difficult to obtain. Also biological NNs are normally complex and nonlinear, but many modelling methods are focused on simplified or linearized NN systems.

40 pt
0.556 in
14.1 mm

Bayesian networks (BN) can represent static dependency among involved variables in a natural way and are intensively applied to multivariable biological network data analysis (Jansen et al., 2003; Yu et al., 2004). Somewhat less established, but perhaps of equal or more importance, are the Dynamic Bayesian Networks (DBNs), which can be used to model stochastic evolution of a set of random variables over time horizon (Hanks and Madigan, 2005; Perrin et al., 2003; Zou and Conzen, 2005), thus being more suitable for dynamical causality analysis of multivariable datasets. The DBN method is essentially a multivariable time-series analysis method with the use of the so-called Minimum Description Length (MDL), and it can simultaneously take into account all dynamical couplings in network causality analysis. DBNs have significant advantages over competing representations such as Granger causality test, which is essentially a linear system analysis method, Kalman filters, which handle only unimodal posterior distributions with linear models, and hidden Markov models (HMMs) (Smyth et al., 1997), whose parameterization grows exponentially with increasing number of state variables.

In this paper, the DBN identification method is investigated to identify underlying dependent structures in multivariable dynamic systems through network structure sorting. The remaining of the paper is organized as follows. The DBN structure identification method is presented in Section 2. Numerical studies are conducted in Section 3. Conclusions and discussions are made in Section 4.

2. DBN STRUCTURE IDENTIFICATION METHOD

2.1 Stage Dynamic Bayesian Networks

Consider a biological network with n nodes that can be represented as a discrete variable set $\mathbf{X}=\{X_1,X_2,\cdots,X_n\}$. Each state variable (or node), X_i ($i=1,\cdots,n$), has s_i possible values: $w_{i1},w_{i2},\ \cdots w_{is_i}$. The parent set of X_i is denoted as $\mathrm{Pa}(X_i)$, which contains q_i nodes and assumes a total number of p_i possible unique instantiations, i.e.,

$$p_i=\prod_{X_i\in \mathrm{Pa}(X_i)} s_i \tag{1}$$

Using v_{ij} ($j=1,2,\cdots,p_i$) to represent the j-th unique instantiation in $\mathrm{Pa}(X_i)$, the conditional probability mass function, $\mathrm{P}\left(X_i=w_{ik}\middle|\mathrm{Pa}(X_i)=v_{ij}\right)=\theta_{i,j,k}$, denotes the possibility of X_i taking the k-th value ($k=1,2,\cdots,s_i$) when $\mathrm{Pa}(X_i)$ presents its j-th unique instantiation. Clearly for any i and j, there is $\sum_{k=1}^{s_i}\theta_{i,j,k}=1$. All values of $\theta_{i,j,k}$ compose the conditional probability table (CPT) for the network.

A BN is defined as a directed acyclic graph (DAG) with a joint probability distribution over a variable set $\mathbf{X}$, denoting formally as $BN=(G,\Theta)$. The vertices of a DAG correspond to the network variables, $\{X_1,X_2,\cdots,X_n\}$; Θ is composed of all entries in the CPT given the network structure of G. It is reasonable to assume that each node X_i is independent of its non-descendants given its parent nodes. Following the chain rule, the joint probability of a BN is the product of the conditional probabilities specified for each variable, i.e.

$$\mathrm{P}(\mathbf{X})=\mathrm{P}(X_1,X_2,...,X_n)=\prod_{i=1}^{n}\mathrm{P}\left(X_i\middle|\mathrm{Pa}(X_i)\right) \tag{2}$$

While a BN is used to describe the probability distribution of a biological network over a static data set, a DBN extends this representation to modelling of dynamical processes. In the formulation of a DBN, the joint probability distribution over the time series, $t=0,1,\cdots,T$, can be written as:

$$\mathrm{P}\left(\mathbf{X}(0),\mathbf{X}(1),\cdots,\mathbf{X}(T)\right)=\mathrm{P}\left(\mathbf{X}(0)\right)\cdot\prod_{t=1}^{T}\mathrm{P}\left(\mathbf{X}(t)\middle|\mathbf{X}(t-1)\right) \tag{3}$$

where $\mathbf{X}(0)$ represents the initial states of the prior network. With the first -order Markov assumption, the state transition probability can be calculated by means of a DAG as follows:

$$\mathrm{P}_{DB}\left(\mathbf{X}(t)\middle|\mathbf{X}(t-1)\right)=\prod_{i=1}^{n}\mathrm{P}_{DB}\left(X_i(t)\middle|\mathrm{Pa}\left(X_i(t)\right)\right) \tag{4}$$

2.2 Minimum Description Length Criteria

There are various methods to evaluate the so-called "loss function" of BNs or DBNs with respect to a certain network data set, among which typically used criteria include Expectation Maximization (EM) (Friedman, 1998), Bayesian Information Criterion (BIC) score (Schwarz, 1978), BDe score (Heckerman et al., 1995), and Minimum Description Length (MDL) score. The MDL algorithm was originally proposed from studies on universal coding (Rissanen, 1978). In order to save/restore a group of given instance data into/from computer memory, a model is usually used to compress the coded data. Thus, the required data length, i.e., the total description length, is equal to the summation of the data length for saving the model parameters and the length of the compressed network response data. The best model is the model with an MDL. This method was later on applied to BNs' structure learning (Lam and Bacchus, 1994).

According to MDL criterion, the optimal BN structure is the network with a minimum summation of the network parameter length and the network response data length. This means that a balance needs to be achieved between the network complexity and the matching degree of the network with the response data. For this purpose, a score of MDL consists of the following two parts:

(i) The network parameter data length

For an n-nodes BN, it requires data with a length of $\log_2(n)$ to code the indexes for each node (using binary coding). The i-th node can be coded by $q_i\log_2(n)$ bits. Therefore, to store a structure of BN, the required binary data length is

$$DL_{struct}(BN)=\log_2(n)\cdot\sum_{i=1}^{n}q_i \tag{5}$$

In the storage of a CPT, each variable X_i has $p_i\times(s_i-1)$ parameters. Since $\sum_{k=1}^{s_i}\theta_{i,j,k}=1$ for any i and j in the network, only (s_i-1) independent parameters (instead of s_i) are required in each column of the CPT. To save each parameter, it requires $\frac{1}{2}\log_2(m)$ bits, where m is the number of data instance of $\mathbf{X}$. For a DBN, the number of m is taken as the length of time series, i.e., $m=T+1$. Thus, the total length required for saving a full CPT of the DBN is

$$DL_{tab}(BN)=\frac{1}{2}\log_2(T+1)\cdot\sum_{i=1}^{n}p_i\cdot(s_i-1) \tag{6}$$

(ii) The compressed length of the response data

Suppose all data instances are independent and the data set (D) is complete, then the binary length of all instance data is

$$\begin{aligned}DL_{data}(D\,|\,BN)&=-\log_2\mathrm{P}(D\,|\,BN)\\&=-\log_2\left(\mathrm{P}\left(\mathbf{X}(0)\right)\cdot\prod_{t=1}^{T}\mathrm{P}\left(\mathbf{X}(t)\middle|\mathbf{X}(t-1)\right)\right)\\&=-\log_2\prod_{i=1}^{n}\prod_{j=1}^{p_i}\prod_{k=1}^{s_i}\left(N_{ijk}/N_{ij}\right)^{N_{ijk}}\end{aligned} \tag{7}$$

where N_{ijk} counts the number of instances when $X_i=w_{ik}$ under the cause of $\mathrm{Pa}(X_i)=v_{ij}$, N_{ij} is the total number of

instances when $\mathrm{Pa}(X_i)=v_{ij}$ for all values of X_i, i.e., $N_{ij}=\sum_{k=1}^{s_i}N_{ijk}$. Taking into account the three data length measures in (5)-(7), a total MDL score of a structured *BN* depending on all data instance is

$$\begin{aligned} &Score_{MDL}(BN:D)\\ &=DL_{struct}(BN)+DL_{tab}(BN)+DL_{data}(D\mid BN)\\ &=\log_2(n)\cdot\sum_{i=1}^{n}q_i+\frac{1}{2}\log_2(T+1)\cdot\sum_{i=1}^{n}p_i\cdot(s_i-1)\\ &-\log_2\prod_{i=1}^{n}\prod_{j=1}^{p_i}\prod_{k=1}^{s_i}\left(N_{ijk}/N_{ij}\right)^{N_{ijk}} \end{aligned} \tag{8}$$

By the MDL criteria, the optimal model structure is selected as the one with the MDL.

2.3 Genetic Algorithm for Network Structure Optimization

Use of efficient searching algorithms is crucial in identifying the optimal structure with the lowest loss function. Currently, the greedy searching or the best first searching are commonly used for BN structure optimization because of their computational simplicity. However, these two algorithms lead to local optimum solutions that are dependent on the initial structure. Alternatively, the simulated annealing method, a global optimization algorithm, can be used, but this algorithm is normally difficult to converge and consumes a large computational time. In this work, we propose to employ another global optimization algorithm, Genetic Algorithm (GA), in structure identification of DBN. Details of a classical GA can be found in (Holland, 1975; Whitley, 2014).

2.4 Pulsed Neural Networks

To test the effectiveness of the DBN identification method for network structure sorting, we construct synthetic NNs using Pulsed Neural Networks (PNN) and calculate their network response data. The PNNs, also called as the third generation of artificial NN, are based on spiking neurons, or "integrate and fire" neurons (Maass and Bishop, 1999). These neurons utilize recent insights from neurophysiology, specifically the use of temporal coding to pass information between neurons (Buracas et al., 1998; Kasabov, 2014; Mainen and Sejnowski, 1995; Riehle et al., 1997; Thorpe et al., 1996; Vaadia et al., 1995), which can closely mimic realistic communications between neurons. Therefore, PNNs are widely applied to study the properties of NNs.

For a spiking neuron i, denoted by X_i, the membrane voltage can be denoted by x_i. The neuron is fired once x_i reaches a threshold level of δ. The moment of the *i*-th neuron crosses the threshold is recorded by a firing time t_i^f. The set of all firing time instances, called spike train, is described by the following set

$$\Phi_i=\left\{t_i^f\,\middle|\,x_i\left(t_i^f\right)=\delta,\quad f=1,2,\ldots\right\}\quad(i=1,2,\cdots,n) \tag{9}$$

where the superscript f is the index for the firing time. There are two main processes that contribute to the value of X_i. The first one is a negative-valued function $\Psi_i\left(t-t_i^f\right)$ indicating an immediate "reset" after each firing time in Φ_i, where t_i^f is the most recent spike time before the current time t. In the biological context, $\Psi_i(\cdot)$, also called the refractoriness function, is used to describe the neuronal refractoriness after a fired spike. The second contribution is from the *i*-th neuron's pre-synaptic neurons, i.e., $\mathrm{Pa}\left(X_i\right)$.

In the following we use notation l to represent the *l*-th pre-synaptic neuron in the parent set, where $l=1,\cdots,q_i$ for $\mathrm{Pa}\left(X_i\right)$. A spike in the pre-synaptic node at time t_l^f increases (or decreases) X_i of post-synaptic neuron i for $t>t_l^f$ by adding a sum of weighted kernel functions, denoted by $\omega_{il}\cdot\varepsilon_{il}(t-t_l^f)$. The sign of ω_{il} is used to indicate the synaptic efficacy, $\omega_{il}>0$ represents excitatory synapses and $\omega_{il}<0$ for inhibitory synapses. The kernel function $\varepsilon_{il}(\cdot)$, also called the postsynaptic potential (PSP) function, describes the response of the *i*-th neuron due to a pre-synaptic potential at t_l^f. The term $\varepsilon_{il}(\cdot)$ can also be regarded as a measure of the combined effect of axon transmission property and membrane transmission property of neurons. Therefore, the value of the *i*-th neuron at current time t is given by a linear superposition of the two contributions,

$$x_i(t)=\sum_{t_i^f\in\Phi_i}\Psi_i(t-t_i^f)+\sum_{t_l^f\in\Phi_l}\sum_{l=1}^{q_i}\omega_{il}\cdot\varepsilon_{il}\left(t-t_l^f\right) \tag{10}$$

The model in (9)-(10) is referred as the Spike Response Model (SRM) (Maass and Bishop, 1999). These two equations, together with the connectivity topology of NNs, form a simple mechanism for simulation of biological NNs. The noise term can be introduced into the SRM by adding a term of stochastic current $I_i^{noise}(t)$. Then (10) is changed to

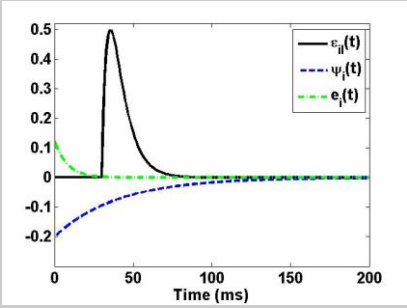
$$\begin{aligned} x_i(t)=&\sum_{t_i^f\in\Phi_i}\Psi_i(t-t_i^f)+\sum_{t_l^f\in\Phi_l}\sum_{l=1}^{q_i}\omega_{il}\cdot\varepsilon_{il}\left(t-t_l^f\right)\\ &+\int_0^{\infty}e_i(s)I_i^{noise}(t-s)\,\mathrm{d}s \end{aligned} \tag{11}$$

Here $e_i(\cdot)$ is the membrane dynamics function, representing the dynamics from the local noise current stimulation to the membrane voltage of the *i*-th neuron.

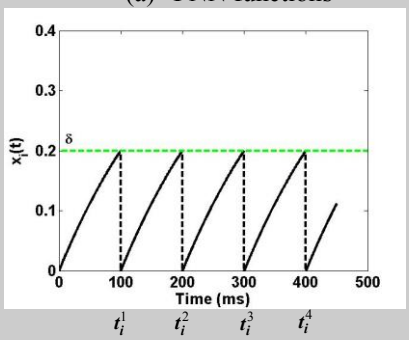
In numerical studies, several typical mathematical formulations can be used to describe the refractoriness function, $\Psi_i(\cdot)$, the PSP function, $\varepsilon_{il}(\cdot)$, and the membrane dynamics function, $e_i(\cdot)$. As an example, the following PSP function can be used

$$\begin{aligned} \varepsilon_{il}(t)=&\frac{1}{1-(\tau_s/\tau_m)}\left[\exp\left(-\frac{t-\Delta_{ax}}{\tau_m}\right)-\exp\left(-\frac{t-\Delta_{ax}}{\tau_s}\right)\right]\\ &\cdot\mathrm{H}(t-\Delta_{ax}) \end{aligned} \tag{12}$$

where τ_s and τ_m are time constants describing axon transmission dynamics and membrane dynamics, respectively; Δ_{ax} is the axonal transmission delay. $\mathrm{H}(t-\Delta_{ax})$ is the Heaviside step function which vanishes for $t \le \Delta_{ax}$ and takes a value of 1 for $t > \Delta_{ax}$.



(a) PNN functions



(b) Membrane voltage of the *i*-th neuron

Fig. 1 Illustration of neuron behaviour with PNN modelling

A typical refractoriness function is

$$\Psi_i(t) = \begin{cases} -\delta \exp\left(-t/\tau\right) & \text{for } t > T^i_{refractory} \\ -\infty & \text{for } t \le T^i_{refractory} \end{cases} \quad (13)$$

where $T^i_{refractory}$ is the absolute refractory period of the *i*-th neuron and τ is the time constant describing refractoriness dynamics. During such a refractory period, a neuron is not fired even if its membrane voltage value goes above the threshold δ. Only when the interval between two successive spikes is greater than $T^i_{refractory}$, the firing activity will be triggered. This well mimics the realistic biological neuron firing mechanism with a refractory period. For the membrane dynamical function, an exponentially damping function is often used to model the first-order inertia dynamics:

$$e_i(t) = \frac{1}{\tau_m}\exp\left(-\frac{t}{\tau_m}\right)\cdot \mathrm{H}(t) \quad (14)$$

Figure 1 illustrates the time profiles of the three functions used in PNN and the time response of x_i. The kernel $\varepsilon_{il}(\cdot)$ describes the response of x_i caused by a pre-synaptic spike at $t=0$; the function $\Psi_i(\cdot)$ reflects the refractoriness after a spike emitted at $t=0$; the the membrane dynamics function $e_i(t)$ represents the dynamics from the local current stimulation to the membrane voltage of a neuron. The parameters are set as follows: $\Delta_{ax} = 30\text{ms}$, $\tau_s = 4\text{ms}$, $\tau_m = 8\text{ms}$, $\delta = 0.2$, and $\tau = 40\text{ms}$. It can be seen that the membrane voltage $x_i(t)$ firing at time t_i^f when it reaches the threshold voltage. After the firing action, it is reset by the function $\Psi_i(\cdot)$ and then re-accumulated by the pre-synaptic spike input $\omega_{il}\cdot\varepsilon_{il}\left(t-t_l^f\right)$.

3. RESULTS

In the following numerical studies, the DBN identification method is applied to PNNs with 2, 3, and 4 nodes, respectively. Details are discussed using a two-node PNN with the following topologies:

(i) A feedback loop, one excitatory connection from neuron 1 to neuron 2 and another reverse excitatory connection of identical strength from neuron 2 to neuron 1 (e.g., $\omega_{21} = \omega_{12} = 0.8$, Fig. 2 row (i));

(ii) An unidirectional connection from neuron 1 to neuron 2 (e.g., $\omega_{21} = 0.8$ and $\omega_{12} = 0$, see Fig. 2 row (ii));

(iii) An unidirectional connection from neuron 2 to neuron 1 (e.g., $\omega_{12} = 0.8$ and $\omega_{21} = 0$, see Fig. 2 row (iii));

(iv) No physical connection between neuron 1 and neuron 2, Fig. 2 row (iv).

For the above four topologies, the identical parameter values $\{3.5\text{ms}, 8\text{ms}, 50\text{ms}, 40\text{ms}, 2\text{ms}, 0.2\}$ was assigned to $\{\tau_s, \tau_m, \Delta_{ax}, \tau, T^i_{refractory}, \delta\}$ for each neuron. Using the DBN model in (11), each neuron is driven by a stochastic noise current $I_i^{noise}(t)$, $(i=1,2)$. We assumed that the noise follows a normal distribution $N(0,1)$ and the noise sources were independent from one another. The spike trains of each network are displayed at a length of 1 second, extracted from a 5s-long simulation (Fig. 2(ii)). During the simulation, the absolute refractory period was set to 2ms. Consequently, all the spike intervals are larger than this refractory period in the raster plot. We binned two spike recordings in a small time interval, $\Delta t = 10$ ms, and then counted the number of spikes in each interval. If the number of the spikes in one interval is equal or greater than 1, let the sample value be "1". Otherwise, if there are not any spikes in the interval, it is valued as "0". The spike trains are sampled according to their differential increments in each time bin Δt, which has the same effect as adding a low-pass filter to the original multivariate time series data. Therefore the high frequency dynamics can be alleviated by applying the moving time window for sampling data. The detailed description of this sampling process can be found in an earlier work (Willie, 1982). With regard to the network sorting, the GA algorithm is employed to search for the optimal structure based on the simulation multivariate time series. The major parameters for GA are selected as follows: the population size is 100, the elite size is 10, the crossover fraction level is 0.8, and the mutation rate is 0.02. The chromosome string is constructed by a binary vector, whose elements come from the adjacent matrix corresponding to the network topology.

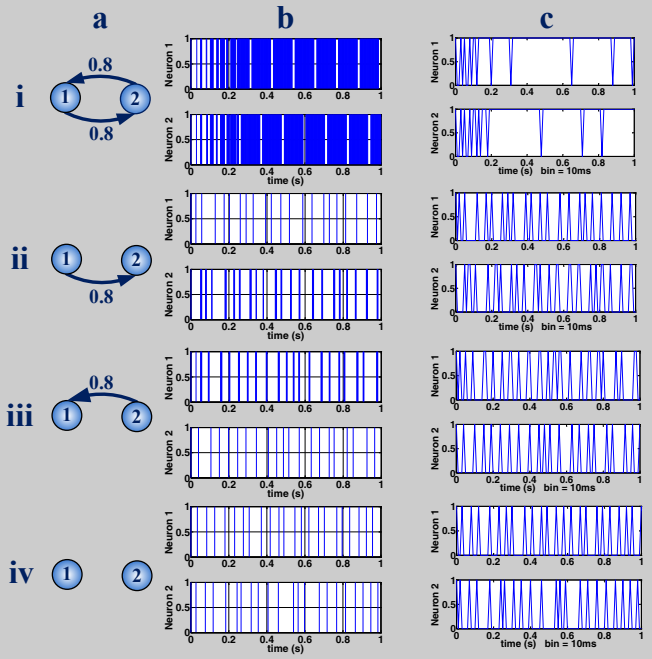


Fig. 2. Simulation of two-node PNNs of different topologies (i)-(iv). Column (a): network topology. The numerical value on each arrow represents the strength of each synaptic interaction. Column (b): the raster plots of two neurons' recordings (the total simulation period is 5s and only 1s is shown here). Column (c): sampled time-series obtained from the two spike trains after binning at 10ms intervals.

Table 1. BDN identification results for 2, 3, and 4-node PNN

Network Scale	# of Structure	# of Connections & Disconnections	Correct Ratio	False Plositve Ratio	False Negative Ratio
2-node	100	200	0.8953	0	0.1047
3-node	100	600	0.9695	0.0047	0.0258
4-node	100	1200	0.9413	0.0104	0.0483

The sampled time series data of the two neurons in each network were created through binning two spike recordings in $\Delta t = 10$ ms and binarizing the number of spikes in each interval as shown in Fig. 2, column (c). Based on these time-series data, the DBN identification method was applied to the four networks (i)-(iv), respectively. As a result (see Table 1, the causal connections can be successfully identified in all 2-node network topologies.

The DBN identification method is also applied to 3-node and 4-node PNN structures. We randomly construct 100 network structures for each network scale and conduct similar simulations as to the 2-node networks. The self-connections of every node are prohibited, thus the total number of connections and disconnections is $n(n-1)$. Key features such as the correct identification ratio, the false positive identification ratio, and the false negative identification ratio are calculated according to the total number of connections and disconnections, and the identification results are also listed in Table 1. It can be seen that a high level of correct ratio has been achieved for all cases. The false positive ratio is low and it increases slightly when the network node number is increased. It is worthwhile noting that the false negative ratio is always greater than the false positive ratio. This fact suggests that the DBN identification method has a stronger tendency of missing existing connections than non-existing connections.

4. CONCLUSIONS AND DISCUSSIONS

In this work, a DBN identification method is developed to investigate biological PNN structure. Benefited from MDL and GA, the DBN identification method is able to rapidly identify and locate causal connections without *a priori* knowledge on the synthetic biological NN architectures. The simulation shows that this method is robust to high levels of measurement noises, and is efficient for nonlinear neuronal dynamics. In addition, the proposed method requires only the measurement of dynamic expression profiles of network nodes. With the advances in high throughput measurement methods, such as multiple single neuron recordings, voltage sensitive dyes or multi-electrode array technology, this modelling technique may soon become applicable in describing large scale *in vitro* or *in vivo* NNs and improve the understanding of relationship between biological dynamics and network topologies.

Other than identifying the functional connections, the DBN identification method also retrieves plenty of connective strength information which is contained in the CPTs. It is observed that a larger value of $\theta_{i,j,k}$ in CPT indeed indicates a more significant regulation to the *i*-th node from its parent node, likewise a smaller coefficient corresponds to a weaker regulation. The quantitative combination in X_i and $\mathrm{Pa}(X_i)$ reflects whether the *i*-th neuron is up-regulated or down-regulated by its parent node, i.e. the property of causal connections. Therefore, such implicit information provides a possibility to deduce more detailed intracellular interaction architectures by collecting the information of the distribution of causality saved in CPTs.

The DBN identification method postulates all neurons accessible for direct measurements. This requirement is, however, not always fulfilled *in vivo*. To overcome this disadvantage, experiments can be designed among those detectable network nodes. This means only partial observations of the network are accessible. The proposed DBN identification method needs to be extended for partial observation neural networks.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (61364018), the Inner Mongolia Autonomous Region Natural Science Foundation (2016 JQ07), the Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region (NJYT-15-A05), and the State Scholarship Fund (201508155034).

REFERENCES

Blinowska, K. J., Kus, R. and Kaminski, M. (2004). Granger causality and information flow in multivariate processes. *Phys Rev E Stat Nonlin Soft Matter Phys,* 70(5), 050902.

Bliss, T. V. P. and Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature,* 361(6407), 31-39.

Buracas, G. T., Zador, A. M., Deweese, M. R. and Albright, T. D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron,* 20(959-969.

Caines, P. and Chan, C. (1975). Feedback between stationary stochastic processes. *IEEE Trans. Autom. Contr.,* 20(4), 498-508.

Doerfler, H., Lyon, D., Nagele, T., Sun, X., Fragner, L., Hadacek, F., Egelhofer, V. and Weckwerth, W. (2013). Granger causality in integrated GC-S and LC-S metabolomics data reveals the interface of primary and secondary metabolism. *Metabolomics,* 9(3), 564-574.

Erickson, J., Tookerb, A., Taib, Y.-C. And Pine, J. (2008). Caged neuron MEA: a system for long-term investigation of cultured neural network connectivity. *J Neurosci Methods*.

Friedman, N. (1998). The Bayesian structural EM algorithm. *Proc. 4th Conf Uncertainty in Artificial Intelligence.* Morgan Kaufmann Publishers Inc., 129-138.

Hanks, S. and Madigan, D. (2005). Probabilistic temporal reasoning. *Foundations of Artificial Intelligence,* 1(315-342.

Heckerman, D., Geiger, D. and Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning,* 20(3), 197-243.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.* U Michigan Press.

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., Snyder, M., Greenblatt, J. F. and Gerstein, M. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science,* 302(5644), 449-453.

Kang, G., Lee, J.-H., Lee, C.-S. and Nam, Y. (2009). Agarose microwell based neuronal micro-circuit arrays on microelectrode arrays for high throughput drug testing. *Lab on a Chip,* 9(22), 3236-3242.

Kasabov, N. K. (2014). NeuCube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks,* 52(62-76.

Kim, S., Kim, J. and Cho, K.-H. (2007). Inferring gene regulatory networks from temporal expression profiles under time-delay and noise. *Comp. Biol. Chem.,* 31(4), 239-245.

Kruger, J. (1983). Simultaneous individual recordings from many cerebral neurons: techniques and results. *Reviews of Physiology, Biochemistry and Pharmacology, Volume 98.* Springer.

Lam, W. and Bacchus, F. (1994). Learning Bayesian belief networks: An approach based on the MDL principle. *Comput. Intell.,* 10(3), 269-293.

Maass, W. and Bishop, C. M. (1999). *Pulsed Neural Networks.* Bradford Book.

Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science,* 268(5216), 1503-1506.

Parsons, T. D., Salzberg, B. M., Obaid, A. L., Raccuia-Behling, F. and Kleinfeld, D. (1991). Long-term optical recording of patterns of electrical activity in ensembles of cultured Aplysia neurons. *J. Neurophysiol.,* 66(1), 316-333.

Perrin, B.-E., Ralaivola, L., Mazurie, A., Bottani, S., Mallet, J. and Dalche-Buc, F. (2003). Gene networks inference using dynamic Bayesian networks. *Bioinformatics,* 19(suppl 2), ii138-ii148.

Riehle, A., Grun, S., Diesmann, M. and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science,* 278(5345), 1950.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica,* 14(5), 465-471.

Schuldiner, S., Shirvan, A. and Linial, M. (1995). Vesicular neurotransmitter transporters: from bacteria to humans. *Physiol Rev,* 75(2), 369-393.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.,* 461-464.

Shmulevich, I., Dougherty, E. R., Kim, S. and Zhang, W. (2002). Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics,* 18(2), 261-274.

Smyth, P., Heckerman, D. and Jordan, M. I. (1997). Probabilistic independence networks for hidden Markov probability models. *Neural Comput.,* 9(2), 227-269.

Spira, M. E. and Hai, A. (2013). Multi-electrode array technologies for neuroscience and cardiology. *Nature Nanotechnology,* 8(2), 83-94.

Thorpe, S., Fize, D. and Marlot, C. (1996). Speed of processing in the human visual system. *Nature,* 381(6582), 520-522.

Trimble, W. S., Linial, M. and Scheller, R. H. (1991). Cellular and molecular biology of the presynaptic nerve terminal. *Annu. Rev. Neurosci.,* 14(1), 93-122.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H. and Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature,* 373(6514), 515--518.

Whitley, D. (2014). An executable model of a simple genetic algorithm. *Foundations of Genetic Algorithms,* 2(1519), 45-62.

Willie, J. S. (1982). Covariation of a time series and a point process. *Jl. Appl. Prob,* 19(609--618.

Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J. and Jarvis, E. D. (2004). Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics,* 20(18), 3594-3603.

Zou, M. and Conzen, S. D. (2005). A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics,* 21(1), 71-79.