2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Complex network from time series

Toshihiro Tanizawa† and Tomomichi Nakamura‡

†Kochi National College of Technology
200-1 Monobe-Otsu, Nankoku, Kochi 783-8508 Japan
‡Graduate School of Simulation Studies, University of Hyogo
7-1-28 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
Email: tanizawa@ee.kochi-ct.ac.jp, tomo@sim.u-hyogo.ac.jp

Abstract—We propose a method of constructing a complex network based on a deterministic model from a time series, in which its time structure is fully and directly incorporated. The network constructed by this method contains time-dependent nodes and links on which positive real numbers are assigned as weight. The method is demonstrated by a known system and shown to be very useful for clarifying hidden time structure embedded in the time series.

1. Introduction

Understanding the complex features of various dynamical systems in the real world continues to be a crucial challenge across the physical and natural sciences. To better understand such complicated interactions it is useful to first transform the system into a new frame of reference. Translating a complicated time-dependent system into a complex network is one of such approaches.

Over the past decade it has become clear that complex networks have a vast range of applicability and give a new perspective on various problems in the real world [1, 2, 3, 4]. We already have several works attempting to combine time series with complex networks, such as networks through the correlation strength [5], recurrence networks [6] and cycle networks for pseudo-periodic time series [7, 8]. These approaches are proven to be effective in understanding complicated and entangled structure of systems [5, 9].

However, these works still do not address one essential ingredient. Although systems in the real world have various different time delay effects in principle, such effects cannot be directly treated or set in the network by current approaches. We refer to the relationship between terms in time series that causes these time delay effects as "time structure." On the other hand, the complex network structures from time series in the existing approaches may not always be sufficient to capture the salient features of the underlying global interrelation structure, because these structures are essentially built on local information such as the temporal structures on an embedding space and correlations (similarities) between each pair of time series among many [5, 6, 9]. In addition, "no similarity" is not equivalent to "no correlation." Even when two signals are not similar, these systems can still have some kind of correlated structures. To handle these, it is crucial to find a method to translate dynamics into the network topology properly.

In this paper we introduce a method of constructing a complex network based on deterministic model structure from a given time series representing a (possibly nonlinear, possibly stochastic) dynamical system. Time structure is directly set in the complex network and the network fully reflects the global structure of interrelation and the hierarchy of the model. Hence, this is the attempt to represent a generic deterministic model as a complex network so as to understand the deterministic dynamics of observed data.

2. Methodology

The method described here is composed of two steps: (i) building a Reduced Auto-Regressive (RAR) model from a given time series and (ii) constructing a complex network from the RAR model.

2.1. Building an RAR model

Given a time series $\{x_t\}_{t=1}^n$ of n observations, an RAR model with the largest time delay w can be expressed by

$$x(t) = a_0+a_1x(t-l_1)+a_2x(t-l_2)+\cdots+a_wx(t-l_w)+\varepsilon(t), \quad (1)$$

where a_i $(i = 0, 1, 2 \ldots, w)$ are unknown parameters, and $\varepsilon(t)$ is assumed to be unknown independent and identically distributed random variables, which are interpreted as fitting errors. The parameters a_i are chosen to minimize the sum of squares of the fitting errors [10].

Among the various information criteria used to find the best (optimal) model, we employ the description length (DL) proposed by Rissanen [11]. The DL formula is

$$\mathrm{DL}(k) = n\ln\frac{\mathbf{e}^T\mathbf{e}}{n} + k\ln n, \quad (2)$$

where n is the number of data points, k is the model size and $\mathbf{e}$ is the fitting errors.

2.2. Constructing a network

After building an RAR model, we transform the model into a directed network (i) by representing each term $x(t)$

at time t by a node labelled by the time and (ii) by drawing an arrow directed from a node $x(t-i)$ to the node $x(t)$, where the time delay term $x(t-i)$ appears in the RAR model for the expression of $x(t)$. This arrow represents the influence of $x(t-i)$ on $x(t)$ with time delay i. We interprete the absolute value of the parameter a_i as the "influence" of $x(t-i)$ on $x(t)$ and transform the influence as a "distance" between nodes $x(t)$ and $x(t-i)$ using a_i on the network space; the larger the absolute value of a_i, the shorter the distance between $x(t)$ and $x(t-i)$. Although there may be several ways to introduce an appropriate "distance," we introduce the following simple distance based on elementary linear algebra [1].

Equation (1) can be interpreted as the scalar product of the coefficient vector

$$\vec{a} \equiv (a_1, a_2, \ldots, a_w), \tag{3}$$

and the set of linearly independent "unit vectors", $(x(t-l_1), x(t-l_2), \ldots, x(t-l_w))$, where the constant parameter a_0 and $\varepsilon(t)$ are not used because these contain no time information. By this interpretation, we introduce the "angle" θ_i between the directions of $x(t)$ and $x(t-i)$ as

$$\theta_i \equiv \arccos\left(\frac{a_i}{\sqrt{a_1^2 + a_2^2 + \cdots + a_w^2}}\right). \tag{4}$$

The distance we introduce should have following properties . Firstly, when vectors $x(t)$ and $x(t-i)$ are in the same direction, the angle θ_i becomes 0 or π and the distance d_i should be 0. We expect the analyticity of the "distance" around $\theta = 0$ and put $d_i \approx \theta_i$ in this case. Secondly, when the vectors $x(t)$ and $x(t-i)$ are perpendicular ($a_i = 0$), the angle θ_i becomes $\pi/2$ and the distance d_i should be infinity. Thus d_i should be inversely proportional to $\cos\theta_i$. Finally, the distance must always be a positive real number. Hence, we define the distance d_i between the nodes $x(t)$ and $x(t-i)$ as

$$d_i \equiv |\tan\theta_i|. \tag{5}$$

According to Eq. (1), the nodes contained in a model are directly connected to $x(t)$. We refer the distance calculated from parameters in the model, Eq. (5), as the direct distance (*DD*). When we take into account time evolution of dynamical system, a pair of nodes can be connected indirectly via some other nodes and we can consider the sum of all distances through the path as the indirect distance (*ID*) between these two nodes. In a network we sometimes find a path with a shorter *ID* than the *DD*. In such a case, we can consider that the indirect path that gives the shortest length is the path on which the information is passed through most effectively and that these two nodes are essentially connected through the shortest *ID* path. We treat the collection of the paths that have the shortest length for any given node pairs as the network of the system. The network constructed in this way reveals the underlying hierarchical structure of the linear model and enables us to know whether the influence of a term may come through other terms.

[1] We understand that the "distance" introduced here depends on the nature of the system, and other distances (e.g. inverse) may be justified in some situations. However, we consider that this distance has broad range of applicability because it reflects the overall balance of the size of parameters in the model. We note that the proposed method is independent of the definition of the "distance"

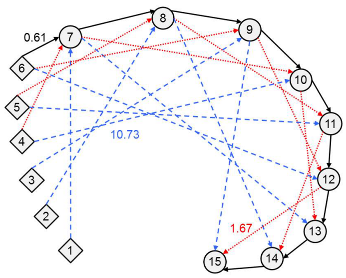

Figure 1: (Color online) Network representation of the first 15 terms (that is, $x(1)$ to $x(15)$) represented by the RAR model, Eq. (6), where the numbers on the nodes are t of $x(t)$ in Eq. (6). The first six nodes represented by squares are the initial nodes. The numbers on the arrows are the direct distances between the nodes. Note that the positions of the nodes are irrelevant (only the topology is important). In this figure, the length of the arrows also does not represent the actual scale.

3. Numerical Example

We demonstrate the application of our algorithm and confirm our theoretical argument with a simple example. We begin by the following RAR model:

$$x(t) = 1.01\ x(t-1) - 0.61\ x(t-3) + 0.11\ x(t-6). \tag{6}$$

By the method we propose here, we obtain

$$\vec{a} = (1.01,\ -0.61,\ 0.11), \tag{7}$$

and using the $\vec{a}$ we introduce a distance between the nodes, direct distances (*DD*s)

$$\vec{d} = (0.6137,\ 1.6655,\ 10.7265). \tag{8}$$

In Fig. 1, we show the overall linkage of the network constructed by our method from Eq. (6) within the time interval from 1 to 15 [2]. The distance between nodes represents the magnitude of influence from the other nodes. According to the model, the nodes $x(t-1)$, $x(t-3)$ and $x(t-6)$ are directly connected with the node $x(t)$. For example, the nodes $x(6)$, $x(4)$ and $x(1)$ are directly connected with

[2] The reason to use the time up to 15 is simply to make the figure simple enough to show the linkage clearly.

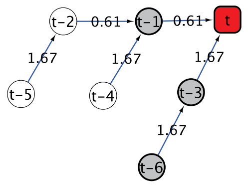

Figure 2: (Color online) The optimal path network constructed from the model, Eq. (6). The nodes shown are from $x(t)$ to the one with the largest time delay, $x(t-6)$. Note that the node $x(t-6)$ is not directly connected to $x(t)$ but connected through $x(t-3)$. The gray color corresponds to terms included in the model and the white color to terms within the largest time delay but not included in the model, respectively. The numbers on the arrows are the direct distances between the nodes. We show all terms within the largest time delay to show the flow of time in the model without gaps and the structure more clearly.

the node $x(7)$ in Fig. 1. However, we note that there are cases where the distance of the directly connected path is not always the shortest. For example, the optimal (shortest) path from $x(9)$ to $x(15)$ is the one via $x(12)$, while the last node $x(15)$ is directly connected with $x(9)$ as Fig. 1 shows. Since both of the direct distances (*DD*s) from $x(9)$ to $x(12)$ and from $x(12)$ to $x(15)$ are 1.67, the indirect distance (*ID*) from $x(9)$ to $x(15)$ is the summation, $1.67 + 1.67 = 3.34$, which is much shorter than the *DD* from $x(9)$ to $x(15)$, 10.73. We can thus conclude that the most significant influence of the term $x(t-6)$ to $x(t)$ is not the direct one but that comes through the term $x(t-3)$.

To show this situation more clearly, we show in Fig. 2 the set of most optimal paths to node $x(t)$ from the nodes within the model (between $x(t)$ and $x(t-6)$). Figure 2 shows that the node $x(t)$ is directly connected with $x(t-1)$ and $x(t-3)$, and the connection from $x(t-6)$ is indirect via $x(t-3)$. This is an outcome of the global structure of interrelation and hierarchy between terms of the model, Eq. (6), and the interplay between the sizes of the parameters and the network topology reveals the time structure. Note that we cannot extract this information by simply examining Eq. (6).

4. Summary

We describe an algorithm to construct complex networks with time structure based on deterministic model structure from time series. In this algorithm a linear model containing various terms of different time delays is transformed into network topology. The advantage of this method over the existing ones is that the global structure of the relationship between terms in a time series model is directly translated to the topology of the corresponding complex network. By extracting the optimal paths for the constructed network, we can find the global structure and hierarchy between terms. In addition, complex networks constructed by the proposed method consists of temporal nodes. Our arguments and computational examples show the effectiveness of introducing complex networks with time structure and that the proposed method has a wide range of applicability and provides us profound insights in investigating time dependent phenomena.

Acknowledgments

One of the authors, T. T., would like to thank the support from the Grant-In-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science, No. 20540382.

References

[1] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, June 1998.

[2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, October 1999.

[3] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[4] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[5] K. Yamasaki, A. Gozolchiani, and S. Havlin. Climate networks around the globe are significantly affected by el niño. *Phys. Rev. Lett*, 100:228501, 2008.

[6] N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths. Complex network approach for recurrence analysis of time series. *Phys. Lett. A*, 373:4246–4254, 2009.

[7] J. Zhang and M. Small. Complex network from pseudoperiodic time series: Topology versus dynamics. *Phys. Rev. Lett*, 96:238701, 2006.

[8] X. Xu, J. Zhang, and M. Small. Superfamily phenomena and motifs of networks induced from time series. *Proceedings of the National Academy of Sciences of the United States of America*, 105:19601–19605, 2008.

[9] M. Small, J. Zhang, and X. Xu. Transforming time series into complex networks. In *Complex Sciences*, pages 2078–2089. Springer Berlin Heidelberg, Shanghai, China, 2009.

[10] K. Judd and A. I. Mees. Embedding as a modeling problem. *Physica D*, 120:273–286, 1998.

[11] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore, 1989.