Physica A 512 (2018) 437–455

Contents lists available at ScienceDirect

Physica A

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

Constructing directed networks from multivariate time series using linear modelling technique

Toshihiro Tanizawa [a,*], Tomomichi Nakamura [b], Fumihiko Taya [c], Michael Small [d,e,f]

[a] Kochi National College of Technology, Monobe-Otsu 200-1, Nankoku, Kochi 783-8508, Japan
[b] Graduate School of Simulation Studies, University of Hyogo, 7-1-28 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
[c] Singapore Institute for Neurotechnology, Centre for Life Sciences, National University of Singapore, 28 Medical Drive, #05-COR, Singapore 117456, Singapore
[d] School of Mathematics and Statistics, The University of Western Australia, 35 Stirling Hwy., Crawley, WA 6009, Australia
[e] Complex Data Modelling Group, The University of Western Australia, 35 Stirling Hwy., Crawley, WA 6009, Australia
[f] Mineral Resources, CSIRO, Kensington, WA 6151, Australia

HIGHLIGHTS

- A method to construct directed networks from multivariate time series is described.
- The method is based on an information theoretic reduction of linear auto-regressive models.
- The method enables to reveal hidden relationships among multivariate time series data.
- The method is generic and can be applied to any kind of multivariate time series data in principle.
- The method is applied to two real time series to verify its effectiveness.

ARTICLE INFO

Article history:
Received 13 April 2017
Received in revised form 9 April 2018
Available online xxxx

Keywords:
Time series modelling
Complex networks
Directed networks

ABSTRACT

We describe a method to construct directed networks from multivariate time series which has several advantages over the widely accepted methods. This method is based on an information theoretic reduction of linear (auto-regressive) models. The models are called reduced auto-regressive (RAR) models. The procedure of the proposed method is composed of three steps: (i) each time series is treated as a basic node of a network, (ii) multivariate RAR models are built and the constituent information in the models is summarized, and (iii) nodes are connected with a directed link based on that summary information. The proposed method is demonstrated for numerical data generated by known systems, and applied to several actual time series of special interest. Although the proposed method can identify connectivity, there are three points to keep in mind: (1) the proposed method cannot always identify nonlinear relationships among components, (2) as constructing RAR models is NP-hard, the network constructed by the proposed method might be near-optimal network when we cannot perform an exhaustive search, and (3) it is difficult to construct appropriate networks when the observational noise is large.

* Corresponding author.
E-mail addresses: tanizawa@ee.kochi-ct.ac.jp (T. Tanizawa), tomo@sim.u-hyogo.ac.jp (T. Nakamura), michael.small@uwa.edu.au (M. Small).

https://doi.org/10.1016/j.physa.2018.08.137
0378-4371/

1. Introduction

Time series of natural phenomena usually show irregular fluctuations, and it is often found that such behaviours can be attributed to time delays (periodicities) in systems [1]. We consider periodicities as an important clue to understand dynamical phenomena in nature, irrespective of whether the data are linear or nonlinear. When data is linear the information of periodicities is directly linked to essential understanding of the linear system generating the data. Nonlinear data might also have periodicities. In this case the periodicities is one of the important clues to understand the underlying characteristic in the data and the system. In this paper, we propose a method to construct directed networks from multivariate time series from the perspective of linear periodic structures (relationships).

There has been various works for constructing networks from multivariate time series and many applications [2–14]. Among them, there are two previously proposed and widely accepted approaches for constructing networks from multivariate time series. One uses information from the frequency domain in which phase differences (shifts or delays) at a frequency between two signals are examined [7], and the other uses information from the time domain in which similarities between two signals are examined in terms of time differences [12].

Although both approaches have proven to be effective in various cases [7,12], we feel that their effectiveness is constrained by the following two possible concerns. The first one is that, as the perspective to construct the networks by these approaches has never been clearly specified, what the constructed networks actually represent is not clear. In these approaches, multivariate auto-regressive (MVAR) model, the cross correlation (CC) function, and a fixed threshold value are applied to investigate the existence of relationship between a pair of time series [7,12]. The second concern is that these statistical approaches often cannot adequately capture more local, nonlinear, or non-stationary peculiarities of time series. As a result, it is not clear what the constructed networks by the methods indicate for the data. Although the perspective of these approaches has never been clearly specified, we consider that the perspective of these approaches corresponds to linear periodic structures, because the MVAR model and the CC function are used. As mentioned above, both approaches are indirect methods to identify underlying linear periodic structures among multivariate time series. We consider that more straightforward approach is preferable to identify subtle features of the structures.

In this paper we propose a method which can construct networks from multivariate time series reflecting their dynamical nature as faithfully as possible based on a firm perspective. The proposed method utilizes a previously proposed linear model, the reduced auto-regressive (RAR) model [15–17]. The RAR model can precisely identify periodicities that are present in a time series, irrespective of whether the data is linear or nonlinear, provided the time series is sufficiently long [17]. Of course, there are restrictions when applying the proposed method. The RAR model cannot always identify nonlinear periodic structure in the data. To build a RAR model we need to find the optimal subset of possible terms for the model, which is expected to be an NP-hard problem. In this case, we usually use a selection algorithm, and the obtained RAR model might be only nearly optimal. It is also difficult to build appropriate RAR models (and to construct appropriate networks) when the observational noise is large.

The paper is organized as follows. We briefly review two widely accepted approaches as the current approaches in Section 2. In Section 3 we identify the network we like to construct from a given set of time series. In Section 4 we describe the problems with the current approaches, and show that the current approaches cannot construct the desired networks. In Sections 5 and 6 we introduce our method and apply the proposed method to several cases using simulated multivariate time series of known linear systems where there are correct linear model systems and the Rössler systems where there is no correct linear system. We discuss difficulties with building RAR models in Section 7. In Section 8 we apply our method to real-world multivariate time series data, which are meteorological data and electroencephalography data.

2. Current approaches

There are two major approaches to construct networks from multivariate time series, which are classified into the frequency-based approach and the time-based approach. In these approaches each individual time series is treated as a basic node of a network and a threshold is used to test the existence of relationship between data.

2.1. Frequency-based approach to network construction

There are also two widely accepted frequency-based methods [7,18,19]. One is Directed Transfer Function (DTF) [2,3], and the other is Partial Directed Coherence (PDC) [4]. DTF was proposed as a multivariate spectral measure to determine the directional influences between any given pair of time series in a multivariate data set [2,3]. DTF is an estimator that simultaneously characterizes the direction and spectral properties of the interaction between signals. PDC was proposed as a factorization of the Partial Coherence after DTF, and PDC is based on MVAR coefficients transformed into the frequency domain [4]. Both methods are based on the MVAR model, and the MVAR model is transformed to the frequency domain by the Fourier transform (or z transformation) to investigate the spectral properties. The pair of nodes corresponding to the chosen two time series is connected with a directed link when a value calculated by both the methods is larger than an appropriately chosen threshold. A threshold value is used to determine whether the values are large enough. See more details on DTF and PDC elsewhere [2–4,7].