Physica D 267 (2014) 49–57

Contents lists available at ScienceDirect

Physica D

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

Causation entropy identifies indirect influences, dominance of neighbors and anticipatory couplings

Jie Sun *, Erik M. Bollt

Department of Mathematics, Clarkson University, Potsdam, NY 13699-5815, United States

HIGHLIGHTS

- We demonstrated the success of transfer entropy in detecting information flow in two oscillators.
- We explored the limitations of transfer entropy for causality inference in various scenarios.
- We developed causation entropy for more reliable inference of causality in networks of coupled oscillators.

ARTICLE INFO

Article history:
Available online 10 July 2013

Keywords:
Causality inference
Causation entropy
Coupled oscillator networks
Blinking couplings

ABSTRACT

Inference of causality is central in nonlinear time series analysis and science in general. A popular approach to infer causality between two processes is to measure the information flow between them in terms of transfer entropy. Using dynamics of coupled oscillator networks, we show that although transfer entropy can successfully detect information flow in two processes, it often results in erroneous identification of network connections under the presence of indirect interactions, dominance of neighbors, or anticipatory couplings. Such effects are found to be profound for time-dependent networks. To overcome these limitations, we develop a measure called *causation entropy* and show that its application can lead to reliable identification of true couplings.

1. Introduction

The long-standing puzzle of "what causes what", formally known as the problem of causality inference, is challenging yet central in science [1,2]. Understanding causal relationship between events has important implications in a wide range of areas including as examples social perception [3], epidemiology [4], and econometrics [5]. It is the reliable inference of causality that allows one to untangle complex causal interactions, make predictions, and ultimately design intervention strategies.

Traditional approach of inferring causality between two stochastic processes is to perform the Granger causality test [6]. A main limitation of this test is that it can only provide information about *linear* dependence between two processes and therefore fails to capture intrinsic nonlinearities that are common in real-world systems. To overcome this difficulty, Schreiber developed the concept of *transfer entropy* between two processes [7]. Transfer entropy measures the uncertainty reduction in inferring the future state of a process by learning the (current and past) states of another process. Being an asymmetric measure by design, transfer entropy is often used to infer the directionality of information flow and further the causality between two processes [8,9]. Recently, it became increasingly popular to use transfer entropy for causality inference in networks of neurons [10,11] and in coupled dynamical systems with parameter mismatches [12], anticipatory couplings [13], and time delays [14]. However, despite the overwhelming number of proposed applications, a clear interpretation of the resulting relationship inferred by transfer entropy is lacking.

In this paper, we study information transfer in the dynamics of small-scale coupled oscillator networks. We show by several examples that causal relationship inferred by transfer entropy is often misleading when the underlying system contains indirect connections, dominance of neighboring dynamics, or anticipatory couplings. To account for these effects, we develop a measure called *causation entropy (CSE)* and show that its appropriate application reveals true coupling structures of the underlying dynamics.

2. Information theory and dynamical systems

In this section we introduce the mathematical tools used in this study, which include elements from both dynamical systems and information theory.

* Corresponding author. Tel.: +1 3152682388.
E-mail addresses: sunj@clarkson.edu (J. Sun), bolltem@clarkson.edu (E.M. Bollt).

0167-2789/$ – see front matter
http://dx.doi.org/10.1016/j.physd.2013.07.001

2.1. Dynamical system as a stochastic process

Our focus of this paper is on discrete dynamical systems of the form

$$x_{t+1} = f(x_t), \quad (1)$$

where $x_t \in \mathcal{D} \subset \mathbb{R}^m$ is the state variable and $f : \mathcal{D} \to \mathcal{D}$ is the dynamic rule of the system. A trajectory (or orbit) $\{x_t\}$ of Eq. (1) naturally represents a time series. For a continuous dynamical system $\dot{x} = f(x)$, a time series can be obtained by sampling its continuous trajectory at discrete time points. The time points are often chosen to spread uniformly in time or to be the time instances at which the trajectory intersects a given manifold that is transversal to the trajectory, called a *Poincaré section* [15].

A natural bridge between dynamical systems and information theory is the formulation of symbolic dynamics, which requires discretization of the phase space. In particular, a finite *topological partition* $P = \{P_1, \ldots, P_m\}$ of the phase space $\mathcal{D}$ is a collection of pairwise disjoint sets in $\mathcal{D}$ whose union is $\mathcal{D}$ [16]. Defining the associated set of symbols $\Omega = \{1, 2, \ldots, m\}$, one can transform a trajectory $\{x_t\}$ into a *symbolic sequence* $\{s_t\}$, where s_t is defined by [17,18]

$$x_t \in P_i \subset \mathcal{D} \Rightarrow s_t = i \in \Omega. \quad (2)$$

Viewing Ω as the sample space, the symbolic sequence $\{s_t\}$ can be seen as a time series of a stochastic process. Define a probability measure over the partition P, as

$$\mu : P \to \mathbb{R}. \quad (3)$$

If μ is *invariant* under the dynamics, then [19,20]

$$\text{Prob}(s_t = i) = \mu(i), \quad \forall\, i \in \Omega,\ t \in \mathbb{R}. \quad (4)$$

A partition P is called a *Markov partition* if it gives rise to a stochastic process that is Markovian, i.e., future states of the process depend only on its current state, and not the past states [21,22].

2.2. Information-theoretical measures: entropy, mutual information and transfer entropy

Consider a discrete random variable X whose probability mass function is denoted by $p(x) = \text{Prob}(X = x)$. To quantify the unpredictability of X, one can calculate its *(information) entropy* defined as

$$H(X) = -\sum_x p(x) \log p(x), \quad (5)$$

where by convention, we use "log" to represent "$\log_2$". In general, $H(X)$ approximates the minimal binary description length L of the random variable X, with the following inequality [21]:

$$H(X) \le L < H(X) + 1. \quad (6)$$

It follows that, among all random variables with c elements, the one with uniform distribution yields the maximum entropy, $\log(c)$.

Consider now two random variables X and Y with joint distribution

$$p(x, y) = \text{Prob}(X = x, Y = y), \quad (7)$$

and conditional distribution

$$p(x|y) = \text{Prob}(X = x|Y = y). \quad (8)$$

The *joint entropy* $H(X, Y)$ and *conditional entropy* $H(X|Y)$ for X and Y are defined, respectively, as

$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y), \quad (9)$$

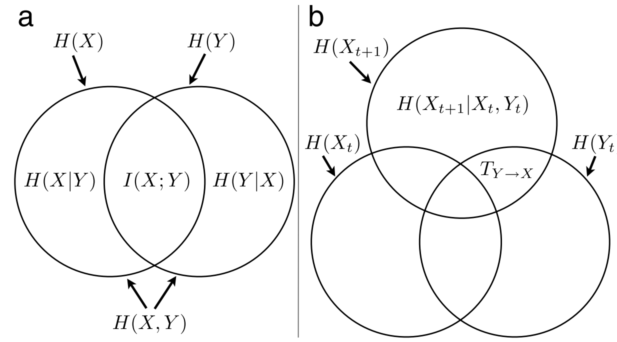

Fig. 1. Venn-like diagrams for information-theoretical measures. (a) Relations between: entropies $H(X)$ and $H(Y)$, joint entropy $H(X, Y)$, conditional entropies $H(X|Y)$ and $H(Y|X)$, and mutual information $I(X; Y)$, of two random variables X and Y. (b) Relations between: transfer entropy $T_{Y\to X}$, entropies of random variables X_{t+1}, X_t, and Y_t, and their joint and conditional entropies. The transfer entropy is the difference between the conditional entropies $H(X_{t+1}|X_t, Y_t)$ and $H(X_{t+1}|X_t)$, which measures the extra information provided by Y_t (in addition to X_t) in the determination of X_{t+1}.

and

$$H(X|Y) = -\sum_y p(y) H(Y|X = x)$$
$$= -\sum_{x,y} p(x, y) \log p(x|y). \quad (10)$$

Similar definition holds for $H(Y|X)$.

It is easy to verify that conditioning reduces entropy, i.e., knowledge of Y will reduce (or at least cannot increase) the uncertainty about X, i.e.,

$$H(X|Y) \le H(X). \quad (11)$$

Similarly, $H(Y|X) \le H(Y)$.

The reduction of uncertainty of X (Y) given full information about Y (X) can be measured by the *mutual information* between X and Y, as [21]

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \quad (12)$$

The mutual information is symmetric in X and Y, and measures their deviation from independence: if X and Y are fully dependent, then $H(X|Y) = H(Y|X) = 0$ and thus $I(X; Y) = H(X) = H(Y)$; on the other hand, if X and Y are independent, then $H(X|Y) = H(X)$ and $H(Y|X) = H(Y)$ and therefore $I(X; Y) = 0$. In general, we have [21]

$$0 \le I(X; Y) \le \min[H(X), H(Y)]. \quad (13)$$

It is convenient to visualize the relationship between entropy, joint entropy, conditional entropy, and mutual information by a Venn-like diagram, as shown in Fig. 1(a).

We now turn to stochastic processes. For a stationary process $\{X_t\}$, its *entropy rate* $H(\mathcal{X})$ can be defined as

$$H(\mathcal{X}) = \lim_{t\to\infty} H(X_t|X_{t-1}, X_{t-2}, \ldots, X_1), \quad (14)$$

which can be thought of as the (asymptotic) growth rate of the joint entropy $H(X_1, X_2, \ldots, X_t)$. If the process is Markovian, then [21]

$$H(\mathcal{X}) = \lim_{t\to\infty} H(X_t|X_{t-1}). \quad (15)$$

For two stochastic processes $\{X_t\}$ and $\{Y_t\}$, the reduction of uncertainty about X_{t+1} due to the information of the past τ_Y states of Y, represented by

$$Y_t^{(\tau_Y)} = (Y_t, Y_{t-1}, \ldots, Y_{t-\tau_Y+1}), \quad (16)$$