Computers and Chemical Engineering 57 (2013) 173–180

Contents lists available at SciVerse ScienceDirect

Computers and Chemical Engineering

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

Data-driven causal inference based on a modified transfer entropy

Yidan Shu, Jinsong Zhao*

State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Received 14 September 2012
Received in revised form 24 March 2013
Accepted 17 May 2013
Available online 25 May 2013

Keywords:
Causal inference
Transfer entropy
Process safety

ABSTRACT

Causality inference and root cause analysis are important for fault diagnosis in the chemical industry. Due to the increasing scale and complexity of chemical processes, data-driven methods become indispensable in causality inference. This paper proposes an approach based on the concept of transfer entropy which was presented by Schreiber in 2000 to generate a causal map. To get a better performance in estimating the time delay of causal relations, a modified form of the transfer entropy is presented in this paper. Case studies on two simulated chemical processes, including the benchmark Tennessee Eastman process are performed to illustrate the effectiveness of this approach.

1. Introduction

Elucidation of the cause-and-effect relationships among variables or events is the central aim of many studies in physical, social, behavioral and biological sciences (Pearl, 2009). In the chemical process industry, knowing the cause-and-effect relationships means knowing the propagation path of a fault or a disturbance, which is critical for alarm management, fault diagnosis, and incident/accident investigations. As a result, it is of great significance to develop an effective and reliable method of causal inference and root cause analysis.

There exist some techniques that are to some extent similar to causal inference and root cause analysis, like HAZOP analysis (Dunjó et al., 2010), and signed digraph (SDG)-based methods (Maurya et al., 2004; Wang et al., 2009; Yang et al., 2010). But methods that rely on only process knowledge are often difficult to use because of the increasing complexity and size of modern industrial processes. Meanwhile, data-driven methods like cross-correlation function (Bauer et al., 2008), and transfer entropy (Schreiber, 2000; Bauer et al., 2007) can overcome such difficulties. However, methods based on statistics of data also have shortages that can lead to ambiguities or false results and fail to discover the real causal structures. So the limitation of data-driven methods needs to be further studied, and combining process knowledge with refined models built by data-driven methods should be considered.

This paper is focused on causal inference based on transfer entropy, and is organized as follows. Section 2 is an introduction of transfer entropy and our modification to transfer entropy. In Section 3, the proposed causal inference approach is introduced. Then case studies are demonstrated in Section 4.

2. Introduction to transfer entropy and proposed modification

2.1. Introduction to transfer entropy

Based on the concept of information theory and information entropy (Shannon & Weaver, 1948), Schreiber proposed the concept of transfer entropy in 2000 to measure the asymmetric interactions in a system. The calculation of transfer entropy is as Eq. (1).

$$t(x|y) = \sum p(x_{i+1}, x_i^{(k)}, y_i^{(l)}) \log \frac{p(x_{i+1}|x_i^{(k)}, y_i^{(l)})}{p(x_{i+1}|x_i^{(k)})} \quad (1)$$

wherein $p(A,B)$ is a joint probability while $p(A/B)$ denotes a conditional probability. x and y represent two variables while x_i and y_i represent their values at time i. $x_i^{(k)} = [x_i, x_{i-1}, \ldots, x_{i-k+1}]$ and $y_i^{(l)} = [y_i, y_{i-1}, \ldots, y_{i-l+1}]$. Transfer entropy represents the difference between the information entropy of x_{i+1} when both $x_i^{(k)}$ and $y_i^{(l)}$ are known and that when only $x_i^{(k)}$ is known. It is to measure the decrease of x's future uncertainty under the condition that y is known.

It should be noted that when Schreiber's definition transfer entropy is based on the assumption that the system can be "approximated by a stationary Markov process", which means that the current state of the system only depends on a certain length of its past. If the Markov property cannot be satisfied, transfer entropy may fail to measure the causal relations in the system.

* Corresponding author.
E-mail address: jinsongzhao@tsinghua.edu.cn (J. Zhao).

0098-1354/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2013.05.011

Unlike mutual information, transfer entropy is in an asymmetric form, which makes it possible to measure cause-and-effect relationships. To consider time delay, which is common in many practical situations, Bauer in 2007 incorporated h, the prediction horizon, and rewrote the transfer entropy as Eq. (2).

$$t(x|y) = \sum_{x_{i+h}, x_i^{(k)}, y_i^{(l)}} p(x_{i+h}, x_i^{(k)}, y_i^{(l)}) \log \frac{p(x_{i+h}|x_i^{(k)}, y_i^{(l)})}{p(x_{i+h}|x_i^{(k)})} \tag{2}$$

The probability density function (PDF) is estimated by a kernel estimator. The kernel function K is centered at every sample point and averaged to estimate the PDF, as shown in Eq. (3).

$$\hat{}p(x) = \frac{1}{N}\sum_{s=1}^{N} K(x - x_s) \tag{3}$$

The Gaussian Kernel function chosen by Bauer is used here:

$$K(x - x_s) = \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{(x - x_s)^2}{2\theta^2}\right) \tag{4}$$

According to Yang's work in 2010,

$$\theta = c \cdot \sigma \cdot N^{-1/5} \tag{5}$$

wherein σ is the standard deviation of the sample points, and $c = (4/3)^{1/5} \approx 1.06$.

For a multivariate case, the joint PDF is:

$$\hat{}p(x_1, x_2, \ldots, x_q) = \frac{1}{N}\sum_{s=1}^{N} K(x_1 - x_{s,1})K(x_2 - x_{s,2}) \ldots K(x_q - x_{s,q}) \tag{6}$$

For each univariate kernel function, parameter θ is calculated by

$$\theta_m = c \cdot \sigma_i \cdot N^{-1/(4+q)}, m = 1, 2, \ldots, q \tag{7}$$

Yang et al. in 2010 applied the Bauer's form of transfer entropy Eq. (2) to validate SDGs. In his work, the prediction horizon was changed to maximize the transfer entropies and the transfer entropies that are large enough will validate the existence of the corresponding SDG arcs that represent the causal relationships between process variables.

2.2. Modification to transfer entropy

With the Bauer's form of transfer entropy that includes the variable of prediction horizon, the prediction horizon that maximizes the transfer entropy may be a good estimation of the time delay of the cause-and-effect. But as what Yang pointed out the time delay inferred with this form of transfer entropy may be inaccurate sometimes.

In our opinion, the reason why Bauer's transfer entropy fails to estimate time delays sometimes is as follows. In Bauer's transfer entropy, the reference of future uncertainty's decrease is $x_i^{(k)}$. As a result, as the prediction horizon varies, the reference also varies. Such a phenomenon seems unreasonable to determine the actual maximized improvement of the prediction.

To solve this problem, a modified transfer entropy form is proposed next. In our modified transfer entropy, the $x_i^{(k)}$ in Eq. (2) is replaced by $x_{i+h-1}^{(k)}$. As a result, the reference is fixed and does not change with h. In our opinion, such a modification will make the maximization more reliable and make it possible to estimate time delay with prediction horizon. The modified form of transfer entropy is as Eq. (8).

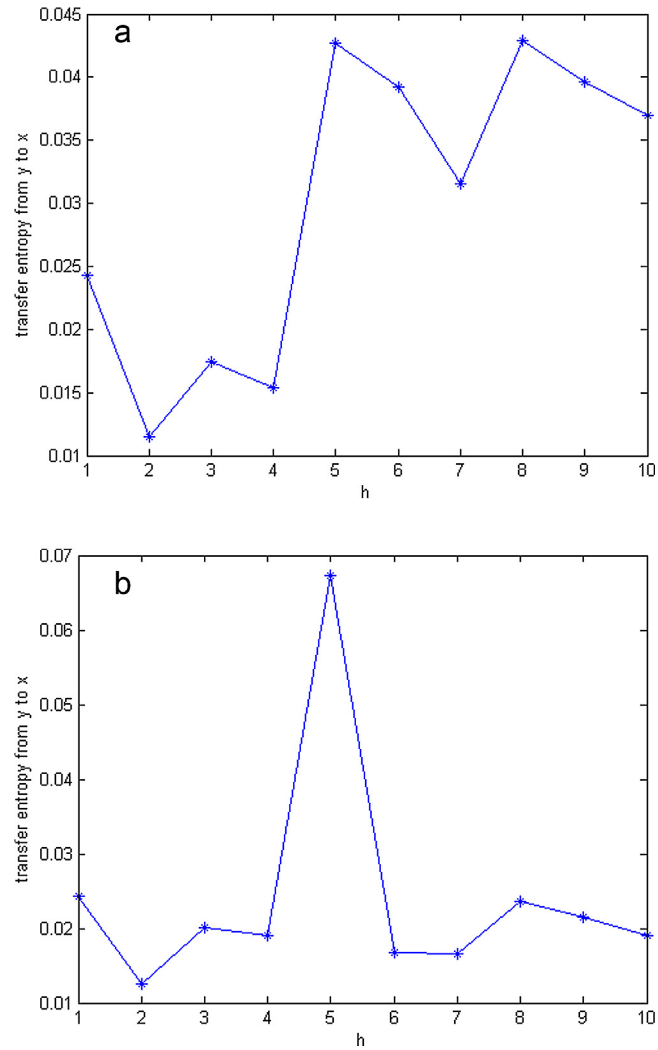

Fig. 1. Transfer entropy vs. *h* under (a) Bauer's form and (b) the modified form.

$$t(x|y) = \sum_{x_{i+h}, x_{i+h-1}^{(k)}, y_i^{(l)}} p(x_{i+h}, x_{i+h-1}^{(k)}, y_i^{(l)}) \log \frac{p(x_{i+h}|x_{i+h-1}^{(k)}, y_i^{(l)})}{p(x_{i+h}|x_{i+h-1}^{(k)})} \tag{8}$$

wherein $x_{i+h-1}^{(k)} = [x_{i+h-1}, x_{i+h-2}, \ldots, x_{i+h-k}]$ and other symbols are the same as Eq. (2).

To illustrate our idea, here is a simple example. Suppose there is an auto-regressive model (see Eq. (9)). y is random and normally distributed at each moment i, while x is determined by both y and x itself:

$$x(i) = x(i-1) + y(i-5) \tag{9}$$

Thus, we can expect the time delay from y to x to be 5.

Both Bauer's form and our modified form of transfer entropy are applied, and let $k = l = 1$ and h be an integer between 1 and 10. Fig. 1(a) and (b) shows transfer entropies as a function of h under both forms.

From Fig. 1(a) we can see Bauer's form can find two local maximums at $h = 5$ and $h = 8$ respectively. Moreover, the latter one is a little bit larger than the former one. That means this form cannot find time delay exactly for this simple example. However, our modified form finds only one maximum at the right h.