ADCONIP 2014 Hiroshima

A New Information Theory-Based Method for Causality Analysis*

Ping Duan[1], Sirish L. Shah[2], Tongwen Chen[1] and Fan Yang[3]

***Abstract*— Detection of causality is an important and challenging problem in root cause and hazard propagation analysis. A new information theory-based measure, transfer 0-entropy, is proposed for causality analysis on the basis of the definitions of 0-entropy and 0-information without assuming a probability space. For the cases of more than two variables, a direct transfer 0-entropy concept is presented to detect whether there is a direct information and/or material flow pathway from one variable to another. Estimation methods for the transfer 0-entropy and the direct transfer 0-entropy are addressed. The effectiveness of the proposed method is illustrated by two numerical examples and one experimental case study.**

I. Introduction

In a large-scale complex system, a simple fault may easily propagate along information and material flow pathways and affect other parts of the system because of the high degree of interconnections among different parts in the system. To determine the root cause(s) of certain abnormality, it is important to capture the process connectivity and find the connecting pathways.

Causality analysis provides an effective way to capture the process connectivity since a causal map can represent the direction of information and/or material propagation pathways [1]. The basic idea of causality can be traced back to Wiener [2] who developed a mathematical definition for causality: Given two random variables X and Y, X could be termed to 'cause' Y if the predictability of Y is improved by incorporating information about X. Granger [3] adapted this definition into the experimental practice, namely, analysis of data observed in consecutive time series. He formalized the prediction idea in the context of linear regression models [3]: X is said to have a causal influence on Y if the variance of the autoregressive prediction error of Y at the present time is reduced by inclusion of past measurements of X. From the definition, we can see that the flow of time is a key point in causality analysis. Therefore, the interaction discovered by causality detection may be unidirectional or bidirectional. This directional interaction is the major difference between causal influence and relations reflected by the symmetric measures such as ordinary coherence and mutual information.

*This work was supported by an NSERC strategic project.

[1]T. Chen and P. Duan are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada tchen@ualberta.ca; pduan@ualberta.ca

[2]S. L. Shah is with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2G6, Canada sirish.shah@ualberta.ca

[3]F. Yang is with the Tsinghua National Laboratory for Information Science and Technology and the Department of Automation, Tsinghua University, Beijing 100084, China yangfan@tsinghua.edu.cn

Inspired by Granger's work, many advanced techniques for causality detection have been proposed, such as the extended and nonlinear Granger causality [4], predictability improvement [5,6], nearest neighbors [7], and transfer entropy (TE) [8–10]. It has been pointed out that TE is a very useful tool in quantifying directional causal influence for both linear and nonlinear relationships [11–13].

TE was initially based on the key concept of Shannon's entropy which is defined stochastically as the averaged number of bits needed to optimally encode a source data set X with the source probability distribution $P(X)$ [14]. Shannon's entropy represents the average unpredictability in a random variable. In other words, it is a measure of the uncertainty associated with a random variable. For a discrete-valued random variable X, assume X has n outcomes $\{x_1, \ldots, x_n\}$, Shannon entropy is defined as [14]

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i),$$

where $p(x_i)$ denotes the probability mass function of the outcome x_i, the base of the logarithm is 2, and the unit is in bits.

One reason for the definition of Shannon's entropy is that random variables in communication systems are generally prone to electronic circuit noises, which obey physical laws yielding well-defined distributions. In contrast, in industrial processes that contain a lot of mechanical and chemical components, the dominant disturbances may not follow a well-defined probability distribution since they may not necessarily arise from circuit noise [15]. Consequently, in process control, disturbances and uncertainties are sometimes treated as bounded unknowns or signals without a priori statistical structure.

One natural question to ask is: without assuming a probability space, is it possible to construct a useful analogue of the stochastic concept of the Shannon's entropy? Hartley entropy or 0-entropy H_0 [16] for discrete variables, and Rényi differential 0th-order entropy or Rényi differential 0-entropy h_0 [17] for continuous variables provide an answer to this question. If a random variable has a known range but an unknown distribution, then its uncertainty can be quantified by the logarithm of the cardinality (H_0) or the logarithm of the Lebesgue measure of its support (h_0). Another natural question is: without assuming a probability space, is it possible to construct a useful analogue of the TE for causality detection? This study is an attempt to provide an answer to this question.

The main contribution of this paper is a new information theory method to detect causal relationships between process variables of linear or non-linear multivariate systems without assuming a probability space. The basic idea of this causality detection method was inspired by the concepts of the 0-entropy and 0-information described in [15] by Nair.

II. Detection of Causality and Direct Causality

In this section, a transfer 0-entropy (T0E) concept based on 0-entropy and 0-information is proposed to detect causality between two variables. In addition to this, direct transfer 0-entropy (DT0E) is proposed to detect whether there is direct causal influence from one variable to another.

A. Preliminaries

Before introducing the concept of the T0E, we describe the non-probabilistic formulations of range, 0-entropy, and 0-information.

A random variable Y can be considered as a mapping from an underlying sample space Ω to a set $\mathbf{Y}$ of interest. Each sample $\omega \in \Omega$ can give rise to a realization $Y(\omega)$ denoted by $y \in \mathbf{Y}$. Then the marginal range of Y is defined as [15]

$$[\![Y]\!] = \{Y(\omega) : \omega \in \Omega\}, \tag{1}$$

where $\{\cdot\}$ indicates a set. Given another random variable X taking values in $\mathbf{X}$, the conditional range of Y given $X(\omega) = x$ is defined as

$$[\![Y|x]\!] = \{Y(\omega) : X(\omega) = x, \omega \in \Omega\}. \tag{2}$$

The relationship between the marginal range of Y and its conditional range given $X(\omega) = x$ satisfies that

$$\bigcup_{x \in [\![X]\!]} [\![Y|x]\!] = [\![Y]\!]. \tag{3}$$

The joint range of Y and X is defined as

$$[\![Y, X]\!] = \{(Y(\omega), X(\omega)) : \omega \in \Omega\}. \tag{4}$$

The joint range is determined by the conditional and marginal ranges as follows:

$$[\![Y, X]\!] = \bigcup_{x \in [\![X]\!]} [\![Y|x]\!] \times \{x\}, \tag{5}$$

where $\times$ represents the Cartesian product.

Variables Y and X are said to be unrelated iff the conditional range satisfies $[\![Y|x]\!] = [\![Y]\!]$, where $x \in [\![X]\!]$. Given another random variable Z taken values in $\mathbf{Z}$, variables Y and X are said to be unrelated conditional on Z iff $[\![Y|x, z]\!] = [\![Y|z]\!]$, where $(x, z) \in [\![X, Z]\!]$ [15].

For example, Fig. 1(a) illustrates the case of two related variables Y and X. For a certain value $x \in [\![X]\!]$, the conditional range $[\![Y|x]\!]$ is strictly contained in the marginal range $[\![Y]\!]$. Note that in this case the joint range $[\![Y, X]\!]$ is also strictly contained in the Cartesian product of marginal ranges, namely, $[\![Y]\!] \times [\![X]\!]$. Fig. 1(b) shows the ranges when Y and X are unrelated. For any $x \in [\![X]\!]$, the conditional range $[\![Y|x]\!]$ coincides with the marginal range $[\![Y]\!]$. Moreover, the joint range $[\![Y, X]\!]$ coincides with $[\![Y]\!] \times [\![X]\!]$.

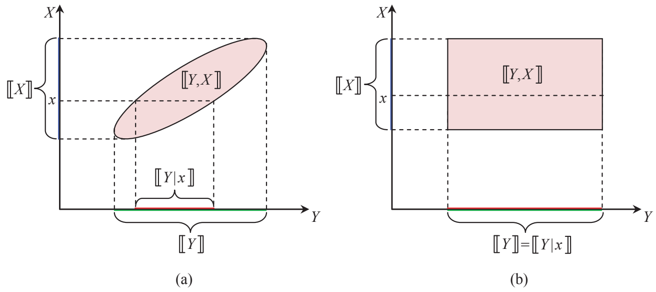

Fig. 1. Examples of marginal, conditional, and joint ranges for related and unrelated random variables (adapted from [15]). (a) Y and X are related; (b) Y and X are unrelated.

Let $|\cdot|$ denote set cardinality and μ denote the Lebesgue measure. A function $\phi([\![Y]\!])$ is defined as

$$\phi([\![Y]\!]) = \begin{cases} |[\![Y]\!]| & \text{for discrete-valued } Y, \\ \mu[\![Y]\!] & \text{for continuous-valued } Y, \end{cases} \tag{6}$$

where $|[\![Y]\!]|$ indicates the set size of $[\![Y]\!]$ for discrete-valued Y, and $\mu[\![Y]\!]$ can be understood as the length of the range $[\![Y]\!]$ for continuous-valued Y. The uncertainty associated with Y can be captured by the (marginal) 0-entropy defined as

$$H_0(Y) = \log \phi([\![Y]\!]), \tag{7}$$

where the base of the logarithm is 2 and the unit of H_0 is in bits. Note that if Y is a discrete-valued random variable, then $H_0(Y)$ represents the (marginal) Hartley entropy or 0-entropy [16] satisfying $H_0(Y) \in [0, \infty)$; if X is continuous-valued, then $H_0(Y)$ indicates the (marginal) Rényi differential 0-entropy [17] which satisfies $H_0(Y) \in (-\infty, \infty)$.

A worst-case approach is taken to define the conditional 0-entropy of Y given X as follows [15,18]:

$$H_0(Y|X) = \text{ess} \sup_{x \in [\![X]\!]} \log \phi([\![Y|x]\!]), \tag{8}$$

where ess $\sup$ represents the essential supremum. $H_0(Y|X)$ can be understood as a measurement of the uncertainty that remains in Y after X is known.

In order to measure the information about Y gained from X, a non-probabilistic 0-information metric, I_0, from X to Y is defined as follows [15,18]:

$$I_0(Y; X) = H_0(Y) - H_0(Y|X) = \text{ess} \inf_{x \in [\![X]\!]} \log(\frac{\phi([\![Y]\!])}{\phi([\![Y|x]\!])}), \tag{9}$$

where ess $\inf$ represents the essential infimum. From the definition, we can see that the 0-information is the worst case log-ratio of the prior to the posterior range set sizes/lengths, and it can be shown that $I_0(Y; X)$ is always non-negative. $I_0(Y; X)$ represents the reduction in uncertainty about Y after X is known; thus, it can be understood as the information about Y provided by X. Note that the definition of the 0-information is asymmetric, that is, $I_0(Y; X) \neq I_0(X; Y)$.

B. Transfer 0-entropy

The concept of 0-information provides an effective way to measure the information about Y provided by X. However, the time flow information is not considered in this definition. Since the time flow information is an important component in causality detection, 0-information cannot be directly used for causality analysis. To incorporate this we propose a transfer 0-entropy concept for causality detection based on the concept of 0-information.

Before introducing the concept of transfer 0-entropy, a conditional 0-information from X to Y given Z is defined as follows:

$$I_0(Y;X|Z) = H_0(Y|Z) - H_0(Y|X,Z), \tag{10}$$

where $H_0(Y|Z)$ and $H_0(Y|X,Z)$ denote conditional 0-entropies defined in (8). The conditional 0-information measures the information about Y provided by X when Z is given.

Now consider two random variables X and Y with marginal ranges $\llbracket X\rrbracket$ and $\llbracket Y\rrbracket$ and joint range $\llbracket X,Y\rrbracket$, let them be sampled at time instant i to get X_i and Y_i with $i = 1,2,\ldots,N$, where N is the number of samples.

Let Y_{i+h} denote the value of Y at time instant $i+h$, that is, h steps in the future from i, and h is referred to as the prediction horizon; $\mathbf{Y}_i^{(k)} = [Y_i, Y_{i-\tau}, \ldots, Y_{i-(k-1)\tau}]$ and $\mathbf{X}_i^{(l)} = [X_i, X_{i-\tau}, \ldots, X_{i-(l-1)\tau}]$ denote embedding vectors with elements from the past values of Y and X, respectively (k and l are the embedding dimensions of Y and X, respectively); τ is the time interval that allows the scaling in time of the embedded vector, which can be set to be $\tau = h$ as a rule of thumb [9]. Let $\mathbf{y}_i^{(k)} = [y_i, y_{i-\tau}, \ldots, y_{i-(k-1)\tau}]$ and $\mathbf{x}_i^{(l)} = [x_i, x_{i-\tau}, \ldots, x_{i-(l-1)\tau}]$ denote realizations of $\mathbf{Y}_i^{(k)}$ and $\mathbf{X}_i^{(l)}$, respectively. Thus $\llbracket Y_{i+h}|\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)}\rrbracket$ denotes the conditional range of Y_{i+h} given $\mathbf{X}_i^{(k)} = \mathbf{x}_i^{(l)}$ and $\mathbf{Y}_i^{(k)} = \mathbf{y}_i^{(k)}$, and $\llbracket Y_{i+h}|\mathbf{y}_i^{(k)}\rrbracket$ denotes the conditional range of Y_{i+h} given $\mathbf{Y}_i^{(k)} = \mathbf{y}_i^{(k)}$. The transfer 0-entropy (T0E) from X to Y is then defined as follows:

$$\begin{aligned}
& T^0_{X\to Y} \\
=\; & I_0(Y_{i+h};\mathbf{X}_i^{(l)}|\mathbf{Y}_i^{(k)}) && (11)\\
=\; & H_0(Y_{i+h}|\mathbf{Y}_i^{(k)}) - H_0(Y_{i+h}|\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)}) && (12)\\
=\; & \operatorname{ess}\sup_{\mathbf{y}_i^{(k)}\in\llbracket\mathbf{Y}_i^{(k)}\rrbracket} \log\phi(\llbracket Y_{i+h}|\mathbf{y}_i^{(k)}\rrbracket) \\
& -\operatorname{ess}\sup_{(\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)})\in\llbracket\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)}\rrbracket} \log\phi(\llbracket Y_{i+h}|\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)}\rrbracket) \\
=\; & \log\frac{\operatorname{ess}\sup_{\mathbf{y}_i^{(k)}\in\llbracket\mathbf{Y}_i^{(k)}\rrbracket}\phi(\llbracket Y_{i+h}|\mathbf{y}_i^{(k)}\rrbracket)}{\operatorname{ess}\sup_{(\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)})\in\llbracket\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)}\rrbracket}\phi(\llbracket Y_{i+h}|\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)}\rrbracket)}, && (13)
\end{aligned}$$

where $\llbracket\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)}\rrbracket$ denotes the joint range of $\mathbf{X}_i^{(k)}$ and $\mathbf{Y}_i^{(k)}$; $\llbracket\mathbf{Y}_i^{(k)}\rrbracket$ denotes the joint range of $\mathbf{Y}_i^{(k)}$.

Since $\bigcup_{\mathbf{x}_i^{(l)}\in\llbracket\mathbf{X}_i^{(l)}\rrbracket}\llbracket Y_{i+h}|\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)}\rrbracket = \llbracket Y_{i+h}|\mathbf{y}_i^{(k)}\rrbracket$, we obtain that $\llbracket Y_{i+h}|\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)}\rrbracket$ is contained in $\llbracket Y_{i+h}|\mathbf{y}_i^{(k)}\rrbracket$; thus, the T0E is always non-negative. From the definition, we can see that the T0E from x to y is the conditional 0-information defined in (10). It measures the information transferred from X to Y given the past information of Y. In other words, the T0E represents the information about a future observation of variable Y obtained from simultaneous observations of past values of both X and Y, after discarding information about the future of Y obtained from past values of X alone. It is obvious that if T0E is greater than zero, then there is causality from X to Y; otherwise, there is no causal influence from X to Y.

From the definition of T0E shown in (13), we can see that the T0E is only related to ranges of the random variables and is independent of their probability distributions. Thus, we do not require a well-defined probability distribution of the data set. This means that the collected sampled data does not need to be stationary, which is a basic assumption for the traditional transfer entropy method.

C. Direct transfer 0-entropy

The T0E measures the amount of information transferred from one variable X to another variable Y. This extracted transfer information represents the total causal influence from X to Y. It is difficult to distinguish whether this influence is along a direct pathway without any intermediate variables or indirect pathways through some intermediate variables [12].

In order to detect whether there is direct causality from X to Y or the causality is indirect through some intermediate variables, a direct transfer 0-entropy (DT0E) from X to Y with intermediate variables $Z_1, Z_2, \ldots, Z_q$ is defined in (14) where $\mathbf{Z}_{j,i_j}^{(s_j)} = [Z_{j,i_j}, Z_{j,i_j-\tau_j}, \ldots, Z_{j,i_j-(s_j-1)\tau_j}]$ denotes the embedding vector with elements from the past values of Z_j for $j = 1,\ldots,q$; $\mathbf{z}_{j,i_j}^{(s_j)}$ denotes a realization of $\mathbf{Z}_{j,i_j}^{(s_j)}$; and $(\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)},\mathbf{z}_{1,i_1}^{(s_1)},\ldots,\mathbf{z}_{q,i_q}^{(s_q)}) \in \llbracket\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)},\mathbf{Z}_{1,i_1}^{(s_1)},\ldots,\mathbf{Z}_{q,i_q}^{(s_q)}\rrbracket$. Note that the intermediate variables are chosen based on calculation results from the T0E [12]. Parameters $s_1,\ldots,s_q$, $i_1,\ldots,i_q$ and $\tau_1,\ldots,\tau_q$ in (14) are determined by the corresponding calculations of the T0E from $Z_1,\ldots,Z_q$ to Y.

III. Calculation Method

In this section, the calculation method for T0E and DT0E is proposed. A method for determination of the confidence levels of T0E is also addressed.

A. Range estimation

From the definition in (13), we can see that a key to T0E estimation is to estimate the joint and conditional ranges. For discrete-valued random variables, joint and conditional ranges can be estimated by finding all possible realizations of the variables. For example, $\llbracket\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)}\rrbracket$ can be obtained by finding all possible realization sets of $(\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)})$; and $\llbracket Y_{i+h}|\mathbf{y}_i^{(k)}\rrbracket$ can be obtained by finding all possible realizations of Y_{i+h} given $\mathbf{Y}_i^{(k)} = \mathbf{y}_i^{(k)}$, and $\phi(\llbracket Y_{i+h}|\mathbf{y}_i^{(k)}\rrbracket)$ is the count of these realizations. For continuous-valued random variables, the estimation of ranges are not as straightforward as the discrete-valued random variables since the realization sets of the continuous-valued random variables are not countable any more. Unfortunately, since most sampled data obtained from industrial processes are continuous-valued, we

$$
\begin{aligned}
D^0_{X\to Y} &= I_0(Y_{i+h};\mathbf{X}_i^{(l)}|\mathbf{Y}_i^{(k)},\mathbf{Z}_{1,i_1}^{(s_1)},\ldots,\mathbf{Z}_{q,i_q}^{(s_q)}) \\
&= H_0(Y_{i+h}|\mathbf{Y}_i^{(k)},\mathbf{Z}_{1,i_1}^{(s_1)},\ldots,\mathbf{Z}_{q,i_q}^{(s_q)}) - H_0(Y_{i+h}|\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)},\mathbf{Z}_{1,i_1}^{(s_1)},\ldots,\mathbf{Z}_{q,i_q}^{(s_q)}) \\
&= \log \frac{\operatorname{ess}\sup_{(\mathbf{y}_i^{(k)},\mathbf{z}_{1,i_1}^{(s_1)},\ldots,\mathbf{z}_{q,i_q}^{(s_q)})} \phi(\llbracket Y_{i+h}|\mathbf{y}_i^{(k)},\mathbf{z}_{1,i_1}^{(s_1)},\ldots,\mathbf{z}_{q,i_q}^{(s_q)}\rrbracket)}{\operatorname{ess}\sup_{(\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)},\mathbf{z}_{1,i_1}^{(s_1)},\ldots,\mathbf{z}_{q,i_q}^{(s_q)})} \phi(\llbracket Y_{i+h}|\mathbf{x}_i^{(l)},\mathbf{y}_i^{(k)},\mathbf{z}_{1,i_1}^{(s_1)},\ldots,\mathbf{z}_{q,i_q}^{(s_q)}\rrbracket)}.
\end{aligned} \tag{14}
$$

need to figure out how to estimate the ranges for continuous-valued variables.

According to (3) and (5), it can be shown that the conditional ranges in (13) are fully determined by the joint range $\llbracket Y_{i+h},\mathbf{X}_i^{(l)},\mathbf{Y}_i^{(k)}\rrbracket$. The joint range can be obtained by the well-developed support estimation method based on the concept of support vector machine (SVM) [19]. Details on the support estimation method can be found in [20–22].

B. Confidence Level Determination of the T0E and DT0E

Small values of the T0E suggest no causality while large values do. The detection of causality can be reformulated as a hypothesis test problem. The null hypothesis is that the T0E measure, $T^0_{X\to Y}$, is small, that is, there is no causality from X to Y. If $T^0_{X\to Y}$ is large, then the null hypothesis can be rejected, which means there is causal influence from X to Y. In order to carry out this hypothesis testing, we may use the Monte Carlo method [9] by constructing a surrogate time series [23]. The constructed surrogate time series must satisfy the null hypothesis that the causal influence from X to Y is completely destroyed; at the same time, the statistical properties of X and Y should not change. In order to construct the surrogate time series that satisfy these two conditions, we propose a new surrogate time series construction method as follows.

Let X and Y be sampled at time instant i and denoted by X_i and Y_i with $i = 1, 2, \ldots, N$, where N is the number of samples; M denotes the length of the training data set, namely, the data size for T0E and DT0E calculations. Then, a pair of surrogate time series for X and Y is constructed as

$$
\begin{cases} X^{\text{surr}} = [X_i, X_{i+1}, \ldots, X_{i+M-1}], \\ Y^{\text{surr}} = [Y_j, X_{j+1}, \ldots, X_{j+M-1}], \end{cases} \tag{15}
$$

where i and j are randomly chosen from $\{1,\ldots,N-M+1\}$ and $\|j-i\| \ge e$, where e is a sufficiently large integer (e is much larger than h) such that there is almost no correlation between X^{surr} and Y^{surr}.

By calculating the T0E from N_s surrogate time series such that $\lambda_n = T^0_{X^{\text{surr}}\to Y^{\text{surr}},n}$ for $n = 1,\ldots,N_s$, the significance level is then defined as

$$
s_{X\to Y} = \frac{T^0_{X\to Y} - \mu_\lambda}{\sigma_\lambda} > 3, \tag{16}
$$

where μ_λ and σ_λ are the mean and standard deviation of λ_n, respectively. Similarly, the value of $s_{X\to Y}$ can also be used as the significance level for the DT0E from X to Y.

IV. Examples and Case Studies

The practicality and utility of the proposed method are illustrated by application to two numerical examples and an experimental data set.

A. Examples

We use simple mathematical equations to represent causal relationships in the following two examples.

Example 1: Assume three linear correlated continuous random variables X, Y, and Z satisfying:

$$
\begin{cases} Y_{k+1} = 0.8X_k + 0.5Y_k + v_{1k} \\ Z_{k+1} = 0.6Y_k + v_{2k}. \end{cases}
$$

where $X_k \sim N(0,1)$; $v_{1k}, v_{2k} \sim N(0,0.1)$; and $Y(0) = 3.2$. The simulation data set consists of 3000 samples. The initial 1000 data points are chosen as the training data and are used for causality analysis.

For ranges estimation, we set $v = 0.01$ since the fraction of outliers of the data is quite small. For determination of γ, the initial 1000 data points are used for training and the remaining 2000 samples are used for validation. Using the cross validation approach, we find that $\gamma = 2^{-2}$ gives good results.

Similar to the TE approach, there are four undetermined parameters in the definition of the T0E in (13): the prediction horizon (h), the time interval (τ), and the embedding dimensions (k and l). The procedure for these parameters determination is similar to that for the TE method proposed in [12]. After the parameters are determined, according to (13) and (16), the T0Es between each pair of X, Y, and Z and the corresponding thresholds (see values within round brackets) obtained via the Monte Carlo method are shown in Table I. Note that the variables listed in column one are the cause variables and the corresponding effect variables appear in the first row. For surrogate time series construction, we set $e = 500$, i.e., $\|j-i\| \ge 500$ in (15), to ensure that there is almost no correlation between each pair of the surrogate data. For the remaining example and case studies, the same value of e is assigned. If the calculated T0E is greater than the corresponding threshold, then we may conclude that the causality is significant; otherwise there is almost no causal influence. Note that if the calculated T0E from one variable to another is zero, then we do not need to calculate the corresponding threshold since it is safe to accept the null hypothesis that there is no causality. From Table I, we can see that X causes Y, Y causes Z, and X causes Z

TABLE I

CALCULATED TRANSFER 0-ENTROPIES AND THRESHOLDS (VALUES IN ROUND BRACKETS) FOR EXAMPLE 1.

$T^0_{\texttt{column 1}\to\texttt{row 1}}$	X	Y	Z
X	NA	1.38 (0.08)	0.60 (0.08)
Y	0.03 (0.07)	NA	1.00 (0.08)
Z	0	0	NA

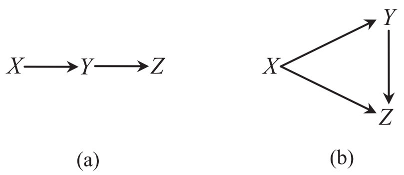

Fig. 2. Information flow pathways for (a) Example 1 and (b) Example 2.

TABLE II

CALCULATED TRANSFER 0-ENTROPIES AND THRESHOLDS (VALUES IN ROUND BRACKETS) FOR EXAMPLE 2.

$T^0_{\texttt{column 1}\to\texttt{row 1}}$	X	Y	Z
X	NA	0.80 (0.07)	0.20 (0.06)
Y	0.03 (0.05)	NA	0.55 (0.08)
Z	0	0	NA

because $T^0_{X\to Y} = 1.38$, $T^0_{Y\to Z} = 1.00$, and $T^0_{X\to Z} = 0.60$ are greater than the threshold. Next we need to determine whether there is direct causality from X to Z. According to (14), we obtain $D^0_{X\to Z} = 0$. Thus, we conclude that there is no direct causality from X to Z. The information flow pathways for Example 1 are shown in Fig. 2(a).

This conclusion is consistent with the mathematical function, from which we can see that there are information flow pathways both from X to Y and from Y to Z, and the information flow from X to Z is indirect through the intermediate variable Y.

Example 2: Assume three nonlinear correlated continuous random variables X, Y, and Z satisfying:

$$\begin{cases} Y_{k+1} = 1 - 2(\mid 0.5 - (0.8X_k + 0.4\sqrt{\mid Y_k \mid}) \mid) + v_{1k} \\ Z_{k+1} = 5(Y_k + 7.2)^2 + 10\sqrt{\mid X_k \mid} + v_{2k}. \end{cases}$$

where $X_k \in [4,5]$ is a uniformly distributed signal; $v_{1k}, v_{2k} \sim N(0, 0.05)$; and $Y(0) = 0.2$. The simulation data consists of 3000 samples. The initial 1000 data points were chosen as the training data and were used for causality analysis.

The T0Es between each pair of X, Y, and Z are shown in Table II. The values within round brackets denote the corresponding thresholds. We may conclude that X causes Y, X causes Z, and Y causes Z because $T^0_{X\to Y} = 0.80$, $T^0_{X\to Z} = 0.20$, and $T^0_{Y\to Z} = 0.55$ are larger than their thresholds.

Thus, we need to first determine whether there is direct causality from X to Z. According to (14), we calculate the DT0E from X to Z with the intermediate variable Y and obtain $D^0_{X\to Z} = 0.23$, which is larger than the threshold 0.06. Thus, we conclude that there is direct causality from X to Z. Secondly, we need to detect whether there is true and direct causality from Y to Z since X is the common source of both Y and Z. We calculate the DT0E from Y to Z with the intermediate variable X and obtain $D^0_{Y\to Z} = 0.39$, which is also larger than the threshold 0.08. Thus, we conclude that there is true and direct causality from Y to Z. The information flow pathways for Example 2 are shown in Fig. 2(b). This conclusion is consistent with the mathematical function, from which we can see that there are direct information flow pathways from X to Y, from X to Z, and from Y to Z.

Compared with the traditional transfer entropy method, the T0E is defined without assuming a statistical space and the only issue is (conditional) ranges of variables. This means that the time series does not need to be stationary, which is a basic assumption for the traditional transfer entropy method. The computational complexity for the T0E estimation is relatively small since we do not need to estimate the joint PDF. Another advantage of the T0E method is that the length of the data does not need to be very large. From the examples described above we can see that 1000 samples are sufficient to give good results while for the transfer entropy estimation the sample number is preferred to be no less than 2000 observations [9].

B. *Experimental case study*

In order to show the effectiveness of the proposed causality detection method for capturing information and/or material flow pathways, a 3-tank experiment was conducted. The schematic of the 3-tank system is shown in Fig. 3. Water is drawn from a reservoir and pumped to tanks 1 and 2 by a gear pump and a three way valve. The water in tank 2 can flow down into tank 3. The water in tanks 1 and 3 eventually flows down into the reservoir. The experiment is conducted under open-loop conditions.

The water levels are measured by level transmitters. We denote the water levels of tanks 1, 2, and 3 by x_1, x_2, and x_3, respectively. The flow rate of the pumped water is measured by a flow meter; we denote this flow rate by x_4. In this experiment, x_4 is set to be a pseudo-random binary sequence (PRBS). The sampled data of 3000 observations were analyzed. Fig. 4 shows the normalized time trends of the measurements. The sampling time is one second. Note that this data set is not strictly stationary since it cannot pass the stationarity test described in [14]. The traditional transfer entropy method may not be suitable for this data set.

The initial 1000 data points are used as training data for γ determination and for causality analysis. The calculated T0Es between each pair of x_1, x_2, x_3, and x_4 are shown in Table III with the thresholds (see values within round brackets) obtained via the Monte Carlo method. If the calculated T0E is larger than the corresponding threshold, then we may conclude that the causality is significant; otherwise there

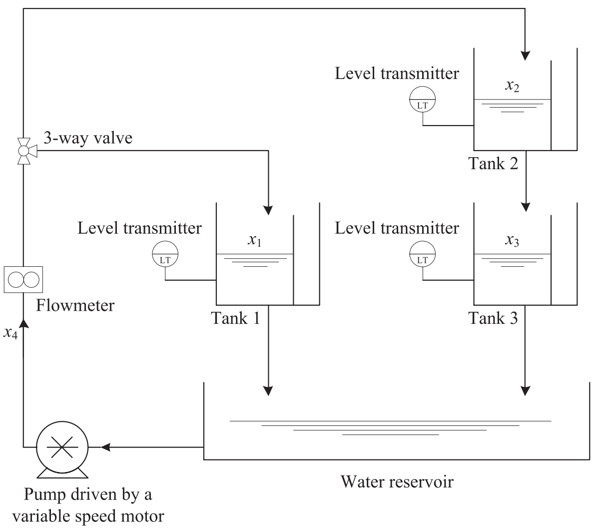

Fig. 3. Schematic of the 3-tank system.

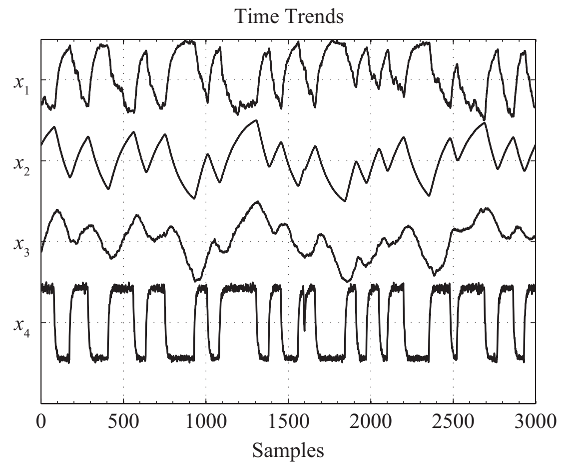

Fig. 4. Time trends of measurements of the 3-tank system.

is no causality. We can see that x_1 and x_2 cause x_3, and x_4 causes x_1, x_2, and x_3. The corresponding connectivity realization is shown in Fig. 5(a).

Now we need to determine whether the causality between x_1 and x_3 and between x_2 and x_3 is true or spurious, as shown in Fig. 5(b). To clarify this we first calculate the DT0E from x_1 to x_3 with intermediate variables x_4 and x_2 and obtain $D^0_{x_1 \to x_3} = 0$, which means that there is no direct information/material flow pathway from x_1 to x_3 and the direct link should be eliminated. Next we calculate the DT0E from x_2 to x_3 with intermediate variable x_4 and obtain $D^0_{x_2 \to x_3} = 0.18$, which is larger than the threshold 0.07. Thus, we conclude that there is true and direct causality from x_2 to x_3. As shown in Fig. 5(b), since x_4 causes x_2, x_2 causes x_3, and x_4 causes x_3, we need to further check whether there is direct causality from x_4 to x_3. According to (14), we calculate the DT0E from x_4 to x_3 with intermediate variable x_2 and obtain $D^0_{x_4 \to x_3} = 0$. Thus, we conclude that there is no direct causality from x_4 to x_3. The corresponding

TABLE III

CALCULATED TRANSFER 0-ENTROPIES AND THRESHOLDS (VALUES IN ROUND BRACKETS) FOR THE 3-TANK SYSTEM.

$T^0_{\texttt{column 1} \to \texttt{row 1}}$	x_1	x_2	x_3	x_4
x_1	NA	0.05 (0.07)	0.14 (0.06)	0.04 (0.06)
x_2	0.05 (0.06)	NA	0.20 (0.07)	0
x_3	0.03 (0.06)	0.04 (0.07)	NA	0
x_4	0.17 (0.06)	0.16 (0.07)	0.06 (0.05)	NA

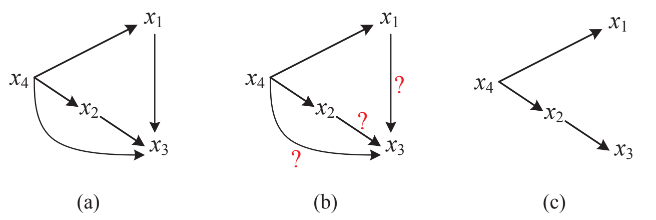

Fig. 5. Information flow pathways for the 3-tank system based on (a) and (b) calculation results of T0Es which represent the total causality including both direct and indirect/spurious causality; (c) calculation results of DT0Es which correctly indicate the direct and true causality.

information flow pathways according to these calculation results are shown in Fig. 5(c), which are consistent with the information and material flow pathways of the physical 3-tank system (see Fig. 3).

C. Industrial case study

Another case study is a part of a flue gas desulfurization (FGD) process at an oil company in Alberta [12]. The schematic of this part of the process is shown in Fig. 6, including a reactor, two tanks, and a pond. Tank 1 receives the overflow from the reactor if the reactor overflows. The liquid in Tank 1 is drawn into the reactor by Pump 1; the liquid in the reactor is drawn into Tank 2 by Pump 2, and the liquid level of the reactor is controlled by adjusting the flow rate of the liquid out of Pump 2; the liquid in Tank 2 is drawn into the pond by Pump 3, and the liquid level of Tank 2 is controlled by adjusting the flow rate of the liquid out of Pump 3. These two level control loops imply that there is bidirectional relationship between the levels and the flow out of the tank due to material as well as information (due to feedback) flow pathways.

We denote the liquid levels of the reactor, Tanks 1 and 2 by y_1, y_2, and y_3, respectively. There is no measurement of the flow rate of the liquid out of Pump 1. We denote the flow rates of the liquid out of Pumps 2 and 3 by y_4 and y_5, respectively. The sampled data of 3544 observations are analyzed. Fig. 7 shows the normalized time trends of the measurements. The sampling time is one minute.

The transfer 0-entropies between each pair of y_1, y_2, y_3, y_4 and y_5 are shown in Table IV with the thresholds (see values within round brackets). The information flow pathways based on the transfer 0-entropies are shown in Fig. 8, we need to further determine whether the causality between y_1, y_2, y_3,

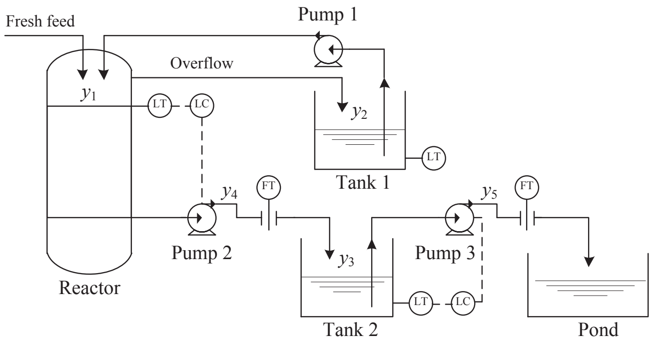

Fig. 6. Schematic of part of the FGD process.

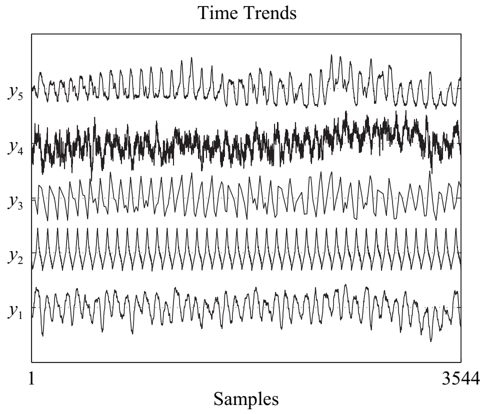

Fig. 7. Time trends of measurements of the FGD process.

TABLE IV

TRANSFER 0-ENTROPIES FOR PART OF FGD PROCESS.

$T^0_{\texttt{column 1}\to\texttt{row 1}}$	y_1	y_2	y_3	y_4	y_5
y_1	NA	0	0.03 (0.06)	0.16 (0.07)	0
y_2	0.15 (0.05)	NA	0.13 (0.06)	0.18 (0.06)	0
y_3	0.12 (0.06)	0	NA	0	0.13 (0.07)
y_4	0.11 (0.05)	0	0.14 (0.06)	NA	0.02 (0.06)
y_5	0	0.04 (0.07)	0.17 (0.06)	0	NA

y_4, and y_5 is true and direct. The calculation results of DT0E are shown in Table V. The information flow pathways based on calculated direct transfer 0-entropies are shown in Fig. 9.

We can see that the connecting pathways shown in Fig. 9 are consistent with the information and material flow pathways of the physical process shown in Fig. 6, where solid lines indicate material flow pathways and dashed lines denote control loops. Note that as mentioned earlier, the bidirectional causality between y_1 and y_4, and between y_3 and y_5 are because of the level feedback control loops.

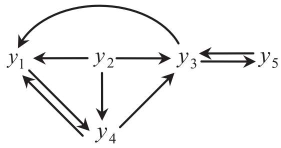

Fig. 8. Information flow pathways for part of FGD process based on calculation results of transfer 0-entropies.

TABLE V

CALCULATED DT0ES FOR PART OF FGD PROCESS.

	Intermediate variable(s)	DT0E
$y_3 \to y_1$	y_2, y_4	0.02
$y_2 \to y_1$	y_4	0.34
$y_2 \to y_4$	y_1	0
$y_2 \to y_3$	y_1, y_4	0.01

V. CONCLUDING REMARKS

A new information theory-based causality detection method based on the T0E has been proposed without assuming a probability space. Moreover, a direct causality detection method based on the DT0E has been presented to detect whether there is a direct information and/or material flow pathway between each pair of variables. The range estimation method for continuous-valued variables and the calculation method for both T0E and DT0E have been addressed. The practicality and utility of the proposed methods have been successfully illustrated by application on two examples and an experimental data set.

The outstanding advantage of the T0E method is that the data does not need to follow a well-defined probability distribution since the T0E is defined without assuming a statistical space and the only issue is its range. This means that the time series does not need to be stationary, which is a basic assumption for the traditional transfer entropy method. This point is clearly illustrated in the experimental 3-tank case study and the industrial case study as presented in Section IV. The analyzed data set is not strictly stationary. The T0E method can still find the information and/or material flow pathways by using this data set. Another advantage of the T0E method compared with the traditional transfer entropy method is that the length of the data does not need to be very large, for example larger than 2000. The reason is that the range estimation is based on the concept of SVM which can handle small sample data sets [24]. Based on our experience, 500 samples are enough to give good results.

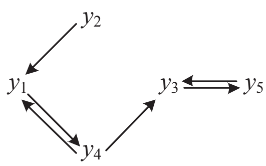

Fig. 9. Information flow pathways for part of FGD process based on calculation results of DT0E.

REFERENCES

[1] F. Yang and D. Xiao, "Progress in root cause and fault propagation analysis of large-scale industrial processes," *Journal of Control Science and Engineering*, vol. 2012, pp. 478 373–1–478 373–10, 2012.

[2] N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers*, E. F. Beckenbach, Ed. New York: McGraw-Hill, 1956.

[3] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[4] N. Ancona, D. Marinazzo, and S. Stramaglia, "Radial basis function approach to nonlinear granger causality of time series," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 70, no. 52, pp. 056 221–1–056 221–7, 2004.

[5] U. Feldmann and J. Bhattacharya, "Predictability improvement as an asymmetrical measure of interdependence in bivariate time series," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 14, no. 2, pp. 505–514, 2004.

[6] M. Bauer, N. F. Thornhill, and J. W. Cox, "Measuring cause and effect between process variables," in *Proceedings of Advanced Process Control Applications for Industry Workshop*, Vancouver, Canada, 2005.

[7] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Nearest neighbors methods for root cause analysis of plantwide disturbances," *Industrial and Engineering Chemistry Research*, vol. 46, no. 18, pp. 5977–5984, 2007.

[8] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, pp. 461–464, 2000.

[9] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Finding the direction of disturbance propagation in a chemical process using transfer entropy," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 1, pp. 12–21, 2007.

[10] P. Duan, S. Shah, T. Chen, and F. Yang, "Methods for root cause diagnosis of plant-wide oscillations," *AIChE Journal*, vol. 60, no. 6, pp. 2019–2034, 2014.

[11] V. A. Vakorin, O. A. Krakovska, and A. R. McIntosh, "Confounding effects of indirect connections on causality estimation," *Journal of Neuroscience Methods*, vol. 184, no. 1, pp. 152–160, 2009.

[12] P. Duan, F. Yang, T. Chen, and S. L. Shah, "Direct causality detection via the transfer entropy approach," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 6, pp. 2052–2066, 2013.

[13] P. Jizba, H. Kleinert, and M. Shefaat, "Rényi's information transfer between financial time series," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 10, pp. 2971–2989, 2012.

[14] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Champaign, Illinois: University of Illinois Press, 1949.

[15] G. N. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1497–1510, 2013.

[16] R. V. L. Hartley, "Transmission of information," *Bell System Technical Journal*, vol. 7, no. 3, pp. 535–563, 1928.

[17] A. Rényi, "On measures of entropy and information," in *Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, USA, 1960, pp. 547–561.

[18] H. Shingin and Y. Ohta, "Disturbance rejection with information constraints: Performance limitations of a scalar system for bounded and gaussian disturbances," *Automatica*, vol. 48, no. 6, pp. 1111–1116, 2012.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[20] B. Schölkopf, J. C. Platt, J. Shawe-taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[21] P. Duan, "Information theory-based approaches for causality analysis with industrial applications," *Ph.D. dissertation, Dept. Electrical and Computer Engineering, University of Alberta, Edmonton, Canada*, 2014.

[22] P. Duan, F. Yang, S. Shah, and T. Chen, "Transfer zero-entropy and its application for capturing cause and effect relationship between variables," *IEEE Transactions on Control Systems Technology*, 2014, submitted for publication.

[23] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, no. 3–4, pp. 346–382, 2000.

[24] L. Gao, Z. Ren, W. Tang, H. Wang, and P. Chen, "Intelligent gearbox diagnosis methods based on svm, wavelet lifting and rbr," *Sensors*, vol. 10, no. 5, pp. 4602–4621, 2010.