Journal of Computers Vol. 33 No. 6, December 2022, pp. 107-117
doi: 10.53106/199115992022123306009

A Prediction Model for Substation Investment Benefit Based on Granger Causality

Xiaoguang Liu[1], Renping Song[1], Guoliang Ding[1], Mingxia Zu[1], Xiaomei Wang[1], Yuan Wang[2*]

[1] Gansu Tongxing Intelligent Technology Development Co., Ltd., China
[2] Northeast Electric Power University, China
166099146@qq.com

Received 9 March 2022; Revised 9 March 2022; Accepted 10 March 2022

Abstract. The construction of a lean operation and inspection integrated management system for substations is an important part of the development and maintenance of the power system. Forecasting the investment benefits of substation project development is an important issue in feasibility analysis. Therefore, we need to use a highly accurate method to make a prediction of the investment benefit of this project. Granger causation is a causal relationship based on "prediction", and inferring about its causality is a key task in time series analysis. In this paper, we propose a new estimation method, Granger causality estimation based on supervised learning. This method uses an eigenvalue representation of the distance between conditional distributions conditioned on past values. And for different time series, the method can give different feature vectors. Applying it to the prediction of the investment efficiency of the substation can achieve a good prediction effect. Therefore, we used granger causality to build a predictive model of the return on investment in substations.

Keywords: Granger Causal relationship, feature vectors, regression model

1 Introduction

With the development of China's economy, the demand for electricity is increasing day by day, and China has carried out a preliminary reform of the electricity market, so investors pay more attention to the economic benefits of investment. At the same time, the project evaluation and research work will effectively promote the standardization and scientificization of the substation project, to further improve the decision-making level of the project. Therefore, we urgently need a forecasting model to predict the benefits of the substation.

Discovering the relationship between cause and effect (Causal relationship) between time dependent variables is one of the important issues in time series analysis, and has broad application prospects. For example, the amount of investment X in research and development (R&D) has an impact on total sales Y, but Y has no impact on X. This causal relationship (X→Y) helps companies make decisions. In addition, discovering causal relationships (control relationships) between genes from time series microarray data is one of the most important tasks in the field of bioinformatics.

Granger as the definition of Causal relationship between time dependent variables, Granger Causal relationship [8] is widely used in various fields [1]. It defines that if the past value of the variable X is useful for predicting the future value of the variable Y, (in the sense of Granger Causal relationship) X is the cause of Y.

The purpose of this study is to establish a Granger Causal relationship estimation method that does not require in depth expertise in data analysis. For this reason, this paper proposes a Granger Causal relationship estimation framework based on supervised learning. Specifically, and the conditional distribution of the future value of Y in consideration of the past value of [x, y] as the distance between these points (maximum mean discrepancy (MMD) [9]) calculate the distance between distributions.

Through experiments, this scheme has higher inference accuracy compared with the existing schemes that use regression models to identify Granger Causal relationship and the existing schemes that infer Causal relationship from i.i.d. data based on classification. In addition, to estimate Granger Causal relationship from multivariate time series data, how to extend the proposal method will also be mentioned. The basic idea of the proposal method is in the literature [6], but the difference between the eigenvectors of grander Causal relationship and different directions, experiments have proved that the results of a kind of artificial data can only show no, and the series of data with multiple variables. There are only actual data in the experiment for the object, and there is not much investigation of the experimental results. In this paper, first, in order to improve the reliability of the experimental results, the effectiveness of the proposed multivariate expansion is verified. In addition, related research on causal

* Corresponding Author

inference methods based on supervised learning and methods for obtaining kernel average estimators will also be detailed.

2 Granger Causal Relationship

As we all know, the existence of correlation between variables does not mean that there is a causal relationship. However, when there is a causal relationship between variables, there is a correlation [2].

In Granger Causal relationship, if the past value of variable X is useful for predicting the future value of variable Y, (in the sense of Granger Causal relationship) X is considered to be the cause of Y. It is defined as follows: Definition 1 (Granger Causal relationship [8]) considers a stable process, that is, a stable random variable sequence $\{X_t, Y_t\}(t \in N)$, but X_t and Y_t are defined on X and Y respectively. Here, S_X and S_Y are set as observations of random variables $\{X_1,\ldots X_t\}$, $\{Y_1,\ldots Y_t\}$, respectively. At this time,

$$P\left(Y_{t+1} \mid S_X, S_Y\right) \neq P\left(Y_{t+1} \mid S_Y\right).$$

If it is true, X_t (in the sense of Granger Causal relationship) is defined as the cause of Y_t.

$$P(Y_{t+1} \mid S_X, S_Y) = P(Y_{t+1} \mid S_Y)\ . \tag{1}$$

If it is true, X_t then (in the sense of Granger Causal relationship) is defined as a cause that is not Y_t.

In order to judge whether the conditional distribution $P(Y_{t+1} \mid S_X, S_Y)$ and $P(Y_{t+1} \mid S_Y)$ are the same or not, in the existing methods [1, 8], according to statistics Hypothesis test to determine whether the conditions attached to the expected value $E[Y_{t+1} \mid S_X, S_Y]$, $E[Y_{t+1} \mid S_Y]$ are equal. This is a much easier question than judging whether equation (1) is true or not. For example, in the existing method [8], these conditional expected values are represented by (V)AR models, and test statistics are calculated based on prediction errors to identify Granger Causal relationship.

When expressing conditional expectations, these methods require appropriate regression models that can explain the data well. However, in fact, it is not easy to choose this regression model. In response to this problem, this paper proposes a new method based on supervised learning.

3 Proposal Method

3.1 Task Settings

In the proposed method, the problem of Granger Causal relationship is solved as a problem of supervised learning. Specifically, the classifier is learned using time series data (training data) with known Causal relationship, and the existence nonexistence and direction of Granger Causal relationship are unknown using the obtained classifier. Solve the problems of supervised learning to estimate the existence nonexistence and the direction of time series data (test data).

Now, suppose that the training data is composed of N pairs of bivariate time series data $S^1,\ldots S^N$. However, [14] suppose that each time S^j series is an observation of random variables $\{\left(X_1^j, Y_1^j\right),\ldots,\left(X_{T_j}^1, Y_{T_j}^1\right)\}$ $(j \in \{1,\ldots,N\})$, where the length is represented by a constant T_j. Here, the individual point series S^j, causal label is called $l^j \in \{+1,-1,0\}$ distribution, which means $Y^j = \left(Y_1^j,\ldots,Y_{T_j}^j\right)$.

Let $v(\cdot)$ be a function that transforms the time series S_j into a single feature vector. In the proposed method, classifier learning is first used. Then, the problem of estimating Granger Causal relationship from the two-variable time series data S (test data) can be said to be the problem of assigning labels to feature vectors using the learned classifier.

As will be described later in Section 3.3, this type of classification task can be extended to multivariate time series data.

3.2 The Design of the Classifier

In order to construct a classifier that assigns a causal label to each time series, we formulate a feature expression v(·). In the following, we will describe how to convert each time series data into feature vectors that are sufficiently different according to the existence nonexistence and direction of Granger Causal relationship between variables.

Design Pointer. In this paper, Granger Causal relationship means, for example, when X is the cause of Y, and Y is not the cause of X, use X→Y to represent the causal label. That is, based on definition 1, the causal label of 3 values is expressed as follows:

$$X \rightarrow Y, if \begin{cases} P(X_{t+1} \mid S_X, S_Y) = P(X_{t+1} \mid S_X) \\ P(Y_{t+1} \mid S_X, S_Y) \neq P(Y_{t+1} \mid S_Y) \end{cases} . \quad \textbf{(2)}$$

$$X \leftarrow Y, if \begin{cases} P(X_{t+1} \mid S_X, S_Y) \neq P(X_{t+1} \mid S_X) \\ P(Y_{t+1} \mid S_X, S_Y) = P(Y_{t+1} \mid S_Y) \end{cases} . \quad \textbf{(3)}$$

$$No\ .\ Caustation\ .\ if \begin{cases} P(X_{t+1} \mid S_X, S_Y) = P(X_{t+1} \mid S_X) \\ P(Y_{t+1} \mid S_X, S_Y) = P(Y_{t+1} \mid S_Y) \end{cases} . \quad \textbf{(4)}$$

To assign a causal label to each time series according to equations (2), (3), (4), it is necessary to determine whether the conditional distribution is the same. To judge whether the conditional distribution is the same, the proposal method uses the kernel average instead of the regression model. This kind of mapping is injective when using characteristic kernels (such as Gaussian kernels), that is, different distributions will not be mapped to the same point [5].

Therefore, if the kernel average is used, the relationship expressed by the equations and inequalities between the conditional distributions in equations (2), (3), and (4) can be expressed as the equations and inequalities between the kernel averages.

The kernel is divided into $P(X_{t+1} \mid S_X, S_Y)$, $P(X_{t+1} \mid S_X)$, $P(Y_{t+1} \mid S_X, S_Y)$, $P(Y_{t+1} \mid S_Y)$, they are respectively mapped into $\mu X_{t+1} \mid S_X, S_Y$, $\mu X_{t+1} \mid S_X \in H_X$, $\mu Y_{t+1} \mid S_X, S_Y$ and $\mu Y_{t+1} \mid S_Y \in H_Y$. Among them, H_X and H_Y are RKHS respectively. At this time, equations (2), (3), and (4) can be rewritten into the following forms.

$$X \rightarrow Y, if \begin{cases} \mu X_{t+1} \mid S_X, S_Y = \mu X_{t+1} \mid S_X \\ \mu Y_{t+1} \mid S_X, S_Y \neq \mu Y_{t+1} \mid S_Y \end{cases} . \quad \textbf{(5)}$$

$$X \leftarrow Y, if \begin{cases} \mu X_{t+1} \mid S_X, S_Y \neq \mu X_{t+1} \mid S_X \\ \mu Y_{t+1} \mid S_X, S_Y = \mu Y_{t+1} \mid S_Y \end{cases} . \quad \textbf{(6)}$$

$$No\ \ Caustation\ .\ f \begin{cases} \mu X_{t+1} \mid S_X, S_Y = \mu X_{t+1} \mid S_X \\ \mu Y_{t+1} \mid S_X, S_Y = \mu Y_{t+1} \mid S_Y \end{cases} . \quad \textbf{(7)}$$

In order to assign the causal label RKHS based on equations (5), (6), (7) whether two points in the RKHS are equal at time t, in other words,the distance between the two points this is what is called in the kernel method community Maximum Mean Difference (MMD) [9] Just judge whether it becomes zero at time t.

In the proposal method, the classification used to estimate Granger Causal relationship uses the feature quantity representation v(.) based on MMD to construct the container. Using MMD, the distance between conditional distributions can be defined as follows [9].

$$MMD^2_{Y_{t+1}} \equiv \| \mu X_{t+1} \mid S_X, S_Y - \mu X_{t+1} \mid S_X \|^2_{H_x} . \quad \textbf{(8)}$$

When estimating the MMD, use the appropriate number of times that can explain the data well, and there is no need to select the regression model, and there is no need to estimate the density function of the conditional distribution. Inspection volume [4] or back check library diversity [11] is more attractive. Because when the former is used to estimate the distance between distributions with conditions, an appropriate regression model should be selected. When the latter is used, the estimation of the density function of the conditional distribution is necessary, because the number of samples in the data is difficult to achieve.

Kernel Average, Estimate of MMD. This Section describes the reason why KKF-CEO can estimate the kernel mean of the distribution conditioned on all past observations by the existing method. This chapter describes how these kernel average estimates are expressed.

For the kernel mean of the conditional distribution $\mu X_{t+1} | S_X, S_Y$, $\mu X_{t+1} | S_X$ is no exception. If KKF-CEO is used, the function ϕx has related components and the inferred form is as follows:

$$\mu X_{t+1} | S_X, S_Y = \sum_{T=2}^{t-1} \omega_T^{XY} \phi x(x_T) . \quad \textbf{(9)}$$

$$\mu X_{t+1} | S_X = \sum_{T=2}^{t-1} \omega_T^{X} \phi x(x_T) . \quad \textbf{(10)}$$

$\omega^{XY} = [\omega_2^{XY}, ..., \omega_{t-1}^{XY}]^T$, $\omega^X = [\omega_2^X, ..., \omega_{t-1}^X]^T (t > 3)$ is a weight vector with real values.

By substituting equations (11) and (12) into equation (8), $MMD^2_{X_{t+1}}$ can be estimated as follows.

$$MMD^2_{X_{t+1}} = \sum_{T=2}^{t-1} \sum_{T'=2}^{t-1} (\omega_T^{XY} \omega_{T'}^{XY} + \omega_T^X \omega_{T'}^X - 2\omega_T^{XY} \omega_{T'}^X) k_x(x_T, x_{T'}) . \quad \textbf{(11)}$$

Kernel Average, Estimate of MMD. To construct a classifier for Granger Causal relationship estimation, the proposed method uses the MMD pair $dt = [MMD^2_{X_{t+1}}, MMD^2_{Y_{t+1}}]^T$ that can be estimated by equation (13) to obtain the feature vector. By using MMD pairs, it can be expected that feature vectors that are sufficiently different for the causal labels of different time series can be obtained. This is because, as can be seen from equations (5), (6) and (7), whether MMD is zero depends on the causal label. In practice, the estimator from a limited data sample is used, so the MMD will not be completely zero, but as shown in Fig. 1, it is expected that the causal label will infer sufficiently different MMD pairs, in fact, in experiments using artificial data, the difference in MMD pairs due to differences in causal labels was confirmed.

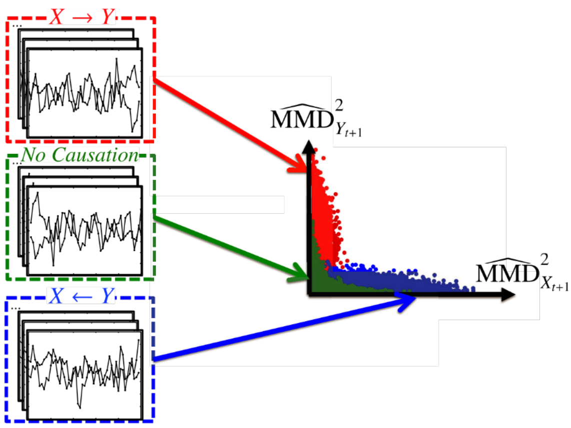

Fig. 1. Transfer function measurement in FRA

Since equations (5), (6) and (7) are independent of time t, this difference in MMD pairs is independent of time t. To take advantage of these differences, the proposed method uses MMD versus dt for each time t. Therefore, from the time series data $S \quad \{(x_1, y_1),...,(x_T, y_T)\}$ of length T, the length is W (W < T) $\{(x_{t-(W-1)}, y_{t-(W-1)}),...,(x_t, y_t)\}$ (t=W, ..., T) ready to use. As a result, we have obtained a series of MMD pairs $\{d_W,...,d_T\}$.

To determine whether the MMD is zero at time t, we use the entire sequence of these MMD pairs to obtain a feature vector. The entire series of MMD pairs can be combined into a feature vector as it is, but if the feature vector is defined in this way, the dimension of the vector becomes 2(T-W+1), and the length T becomes a form, which is a different time The sequence provides feature vectors with different dimensions.

Therefore, in the proposal method, to obtain the feature vector whose number of dimensions does not depend on the time series length T, the kernel average value of the distribution followed by the MMD sequence is considered, and the feature vector is calculated based on the estimator. For this reason, a new kernel k_D, which is different from k_X and k_Y .The unbiased estimator of the kernel of the distribution average Q(D_t),the feature map defined by the kernel k_D on D, $\phi_D(dt) \equiv k_D(dt,\cdot)$ uses the expression feature definition to be the following form.

$$v(S) \equiv \frac{1}{T-W+1}\sum_{t=W}^{T}\phi_D(dt) \ . \tag{12}$$

$$where \quad dt = [MMD^2_{X_{t+1}}, MMD^2_{Y_{t+1}}]^T$$

In formula (14), the feature map $\phi_D(\cdot)$ is the calculation, and the existing method random fourier features (RRF) uses the feature map to randomly sample the low dimensional feature vector of the Fourier transform of the kernel function. As an approximation. In the experiment, the number of features m is set to m = 100, even if a larger m is used for experiments, the estimation accuracy is not greatly improved.

3.3 Extension to Multivariate Time Series

Finally, we describe a method to extend the proposed method to n variable time series ($n \geq 3$).

The Case of 3 Variable Time Series. The feature representation of the 3 variable time series is designed based on conditional Granger Causal relationship [7]. Different from Definition 1, this is the definition of Granger Causal relationship and can be applied to multivariate time series.

According to Definition 1, Granger Causal relationship is estimated from a 3 variable time series, knowing that the wrong result is true. For example, (in the meaning of Granger Causal relationship) there is no causal relationship between variables X and Y, and when the third variable Z is the common cause of X and Y, it is erroneously deduced to be the cause of Y or the cause of X. It is known Can be determined. This is due to the influence of Z, $P(Y_{t+1} | S_X, S_Y) \neq P(Y_{t+1} | S_Y)$ or $P(Y_{t+1} | S_X, S_Y) \neq P(X_{t+1} | S_Y)$ may hold, in addition, There is no Granger Causal relationship between variables X and Y. For example, the situation where X affects Y through Z (i.e., X→Z, Z→Y) is also the case, due to the influence of Z, $P(Y_{t+1} | S_X, S_Y) \neq P(Y_{t+1} | S_Y)$ is sometimes true. Through the same reason, we can know that there is a grander Causal relationship from X to Y is a wrong judgment.

In order to consider the influence of variable z, in conditional grander Causal relationship, the observation S_Z of the defined probability variable $\{Z_1,...,Z_t\}$ on z considers the two conditional distributions of conditional measurement, if $P(Y_{t+1} | S_X, S_Y, S_Z) \neq P(Y_{t+1} | S_Y, S_Z)$ holds, then X is the cause of Y when Z is given, otherwise X is not the cause of Y when Z is given.

In the retrieval method, the introduction of causal label based on this conditional Granger Causal relationship is considered. Same as formula (2), the causal label X → Y

$$X \to Y, if \begin{cases} P(X_{t+1} | S_X, S_Y, S_Z) = P(X_{t+1} | S_X, S_Z) \\ P(Y_{t+1} | S_X, S_Y, S_Z) \neq P(Y_{t+1} | S_Y, S_Z) \end{cases}$$

In short, this can be expressed as follows.

$$X \to Y, if \begin{cases} \mu x_{t+1} \mid s_X, s_Y, s_Z = \mu x_{t+1} \mid s_X, s_Z \\ \mu y_{t+1} \mid s_X, s_Y, s_Z = \mu y_{t+1} \mid s_Y, s_Z \end{cases}.$$

In order to correspond to the situation where there are variables with common causes between two variables, $\mu x_{t+1} \mid s_X, s_Y, s_Z$, $\mu x_{t+1} \mid s_X, s_Z$, $MMD^2{}_{X_{t+1}|Z}$ of MMD between, and $\mu y_{t+1} \mid s_X, s_Y, s_Z$, $\mu y_{t+1} \mid s_Y, s_Z$, $MMD^2{}_{Y_{t+1}|Z}$ of the MMD between are considered to be added to the feature quantity performance. That is, the feature quantity representation (14) is expanded by transforming dt as follows.

$$v(S) \equiv \frac{1}{T-W+1}\sum_{t=W}^{T}\phi_D(dt)\,. \qquad \mathbf{(13)}$$

$$where \quad dt = [MMD^2_{X_{t+1}}, MMD^2_{Y_{t+1}}, MMD^2_{X_{t+1}|Z}, MMD^2_{Y_{t+1}|Z}]^T$$

The Case of n Variable Time Series (n > 3). For the feature quantity representation of the 3 variable time series shown in formula (15), by further adding MMD to dt, it is possible to expand the feature quantity performance of the n variable time series (n > 3), but the n variable in the case of each variable, the number of combinations of variables that may become a common cause in the pairing explodes, so the training data is composed of n variables. Due to the difficulty of adequate preparation, in the proposal method, even in the case of n variables, the feature quantity expression of equation (15) is used.

In the scheme method, for the variable of the ${}_nC_2$ variable formed by selecting two of the n variables, the feature quantity expression of equation (15) is used between each variable pair to estimate the presence or absence of Granger Causal relationship.

Hereinafter, the variable of a certain variable among the n variables is set to X, Y, and the method of estimating the causal relationship between X and Y is described. First, set the three groups of variables (X, Y, Z_v) as $v \in \{1, \ldots, n-2\}$ to consider separately, and use equation (15) to calculate according to the time series data related to these three variables Feature vector. Next, use the learned classifier to calculate the distribution probability of causal label (X → Y, X ← Y, and No Causation) according to each feature vector. Finally, the causal label with the highest allocation probability is allocated in chronological order to infer the Causal relationship.

4 Experiment

4.1 Experimental Setup

The performance of the proposal method (hereinafter referred to as Supervised Inference of Granger Causality (SIGC)), the existing method of inferring Causal relationship through i.i.d data supervised learning, RCC, VAR model, GAM, using Carl regression to identify Granger Causal relationship. The existing methods GC_{VAR} [8], GC_{GAM} , GC_{KER} , and the estimation of Causal relationship based on density function estimation instead of regression model, transfer entropy TE, were compared.

In the proposal method, random forest is used as the classifier. Random forest is selected here because RCC, which is one of the comparison methods, is experimentally higher when using random forest than when using SVM. The estimation accuracy of [9]. In order to prepare the feature vector, the kernel functions k_X , k_Y , and the Gaussian kernel as k_D . The parameter W of the proposed method and the parameter of the existing method have the best performance in each method in the human data experiment described later. As a result, W in the scheme becomes W=12.

4.2 Experiments using 2 Variable Time Series

Learning Classifier. Using 2 variable time series data, learned a classifier for estimating Granger Causal relationship. Same as the existing methods based on supervised learning [3, 10, 13], artificial experimental data and actual data experiments, using artificial data to learn the classifier. This is because there are very few actual data on known Causal relationship.

As training data, 15,000 pages of 2 variable time series data with a length of T = 42 is prepared. Specifically, as shown below, linear time series data and non linear time series data, are prepared.

- Linear time series: The sample comes from the following VAR model.

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \frac{1}{p}\sum_{T=1}^{P} A_T \begin{bmatrix} X_{t-T} \\ Y_{t-T} \end{bmatrix} + \begin{bmatrix} E_{X_t} \\ E_{Y_t} \end{bmatrix}. \quad (14)$$

When the label obtains the time series of X→Y, the coefficient matrix is expressed as

$$A_T = \begin{bmatrix} a_T & 0.0 \\ c_T & b_T \end{bmatrix}.$$

- Non linear time series: Same as above, based on the VAR model, using the standard sigmoid function $g(x) = 1/(1+\exp(-x))$. Sampling. For example, when the series with the label X→Y is obtained, when Y_t depends on $\{[g(X_{t-T}), Y_{t-T}]^T\}_{T=1}^{P}$, X_t only depends on the form $\{X_{t-T}\}_{T=1}^{P}$ Medium sampling.

Manual Data Experiment. Use the linear test data and non linear test data generated as follows to conduct an evaluation experiment.

- Linear test data: A 300 page linear time series is generated based on equation (16). Here, the number of time series with labels X→Y, X→Y and No Causation is set to 100, and the generation of some parameter settings is different from the training data (e.g., the dispersion of noise is given p∈ {0.5, 1.0, 1.5, 2.0}.
- Non linear test data: 300 pages of non linear time series generated, so that the number and label time series X→Y, X→Y has no causal relationship is 100. Here,the nonlinear time series labeled X→Y is generated by the following equation.

$$X_t = 0.2X_{t-1} + 0.9E_{X_t}. \quad (15)$$

$$Y_t = -0.5 + \exp(-(X_{t-1} + X_{t-2})^2) + 0.7\cos(Y^2_{t-1}) + 0.3E_{Y_t}. \quad (16)$$

Here, the noise variables E_{X_t}, E_{Y_t} are sampled from the standard normal distribution N(0, 1). Similarly, a time series with the label X←Y is generated. Prepare the time series with the label No Causation by ignoring the exponential function term in equation (18).

The performance of the proposed method is compared with the existing methods using linear test data and non linear test data. Fig. 2 shows the estimation accuracy of each method. Here, since the proposed method and RCC use randomly generated training data, the average and standard of 20 experiments using different training data.

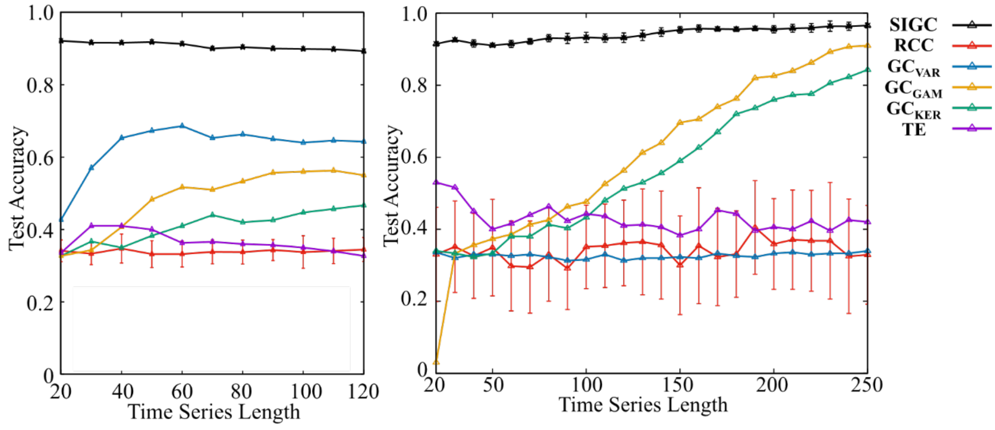

Fig. 2. Test accuracies (left: linear test data; right: nonlinear test data) means and standard deviations (error bars) are shown for our method and RCC based on 20 runs with different training data

First use the non linear test data and use the MMD pair {dt} as a histogram. As mentioned earlier, these MMD pairs are used to calculate feature vectors for each time series.The same experiment was carried out using linear test data,and as shown in Fig. 4, the results suggesting the validity of the feature quantity expression were also obtained.

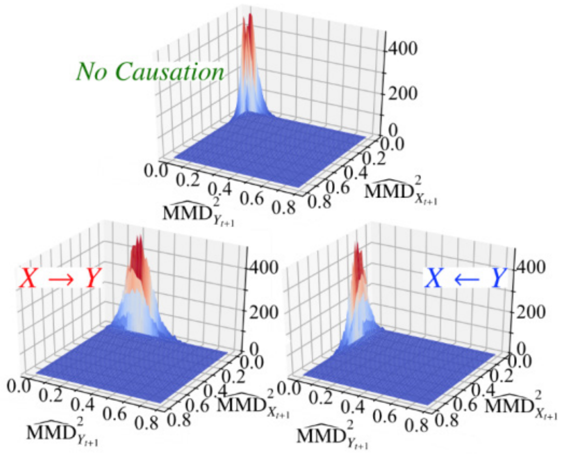

Fig. 3. Histogram of MMDs used to compute the feature vector for each time series in nonlinear test data with X→Y (bottom left), X←Y (bottom right), and No Causation (top)

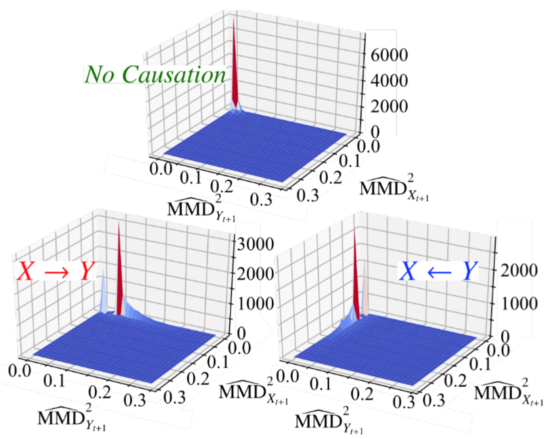

Fig. 4. Histogram of MMDs used to compute the feature vector for each time series in linear test data with X→Y (bottom left), X←Y (bottom right), and No Causation (top)

If completely different MMD pairs can be obtained through causal label, causal label can be assigned even if training data is not used. First, the system used by MMD column, whether the average value of $MMD^2_{X_{t+1}}$ is zero, in order to judge whether the average value of $MMD^2_{Y_{t+1}}$. if two p are used Value and a certain threshold (significance level), a causal label (X→Y, X←Y, or No Causation) can be assigned to each time series.

Actual Data Experiment. The performance of the proposed method is evaluated using actual data. Here, in order to improve the reliability of the experiment, the following two test data sets are prepared:

- The first test data set consists of 5 pairs of time series data downloaded from the database Cause Effect Pairs [11], which is a set of data with known true Causal relationship. For example, River Runoff is a bivariate time series data about average precipitation X and average river discharge Y. The true Causal relationship is called X→Y [13].
- The second test data set is to prepare the above mentioned five pairs of time series by obtaining the time series of each part. Therefore, divide each time series and prepare multiple partial time series with length T = 200.

Table 1 shows the experimental results when using each test data set. Here, the RCC results in Table 1 are omitted. RCC used different training data for 20 experiments, and each time there was a speculation result with completely different output. It can be seen from Table 1 that no matter what the length T of the time series is, the proposed method achieves a higher estimation accuracy than other existing methods.

Table 1. Causal relationships inferred from the first test dataset
('yes' and 'no' denote correct and incorrect results, respectively)

	SIGC	C_{VAR}	GC_{GAM}	GC_{KER}	*TE*
River Runoff (T = 432)	yes	yes	yes	no	yes
Temperature (T = 16382)	yes	no	yes	yes	no
Radiation (T = 8401)	yes	yes	yes	yes	yes
Internet (T = 498)	yes	yes	no	no	yes
Sun Spots (T = 1632)	yes	no	no	no	yes

To confirm that the feature representation returns sufficiently different feature vectors based on the causal label, the same verification experiment was performed in Fig. 3 and Fig. 4.

Specifically, for the 2 variable time series data included in the second test data set, the MMD pair {dt}obtained from each time series is visualized as a histogram, and the result is shown in Fig. 5. The left side of Fig. 5 is the results related to the time series data obtained by River Runoff, Temperature, Radiation whose true Causal relationship is X→Y, and the right side of Fig. 5 is the Internet with real Causal relationship X→Y, obtained by Sun Spots Results related to time series data.

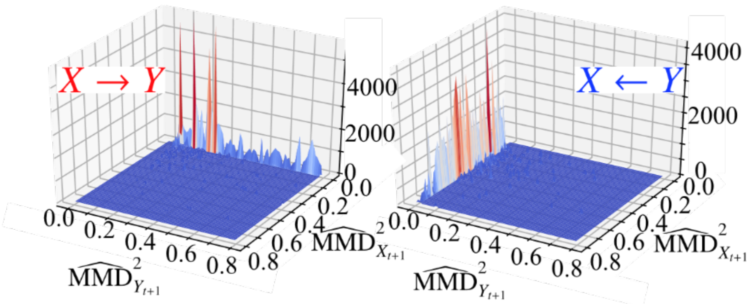

Fig. 5. Histogram of MMDs used to compute the feature vector for each time series in the second test dataset with X→Y (left) and X←Y (right)

It can be seen from Fig. 5 that even with the actual data mentioned above, a completely different MMD is obtained according to causal label.

4.3 Experiments using Multivariate Time Series

The performance of the proposed method was evaluated in a series of data with the number of variables n ≥ 3. The training data were used artificial data and consisted of generated time series data for 3 variables.

Manual Data Experiment. First, use nonlinear artificial data with 3 variables to evaluate the performance of the proposed method. The test data is generated based on the three logistic maps expressed by the following equation.

$$\begin{aligned} X_t &= 0.8(1 - aX_{t-1}^2) + 0.2(1 - aY_{t-1}^2) + sE_{X_t} \\ Y_t &= 1 - aY_{t-1}^2 + sE_{Y_t} \\ Z_t &= 0.8(1 - aZ_{t-1}^2) + 0.2(1 - aX_{t-1}^2) + sE_{Z_t} \end{aligned} \quad . \tag{17}$$

Here, a = 1.8, s = 0.01, and the noise variables E_{X_t}, E_{Y_t}, and E_{Z_t} are samples of the standard normal distribution N(0, 1). The initial values X_1, Y_1, Z_1 are sampled from the uniform distribution U(0, 1), 100 types are prepared, the time series length is T = 1,000, and 100 types of three variable time series data are prepared. It can be seen from equation (19) that the true causal label between the variable pairs (X, Y) and (Z, X) is Y→X, X→Z, and there is no Granger Causal relationship (i.e., No Causation).

Actual Data Experiment. The performance of the proposed method is evaluated using actual data. The following time series microarray data were used as test data.

In this experiment, the number of noncausal gene pairs is much greater than the number of causal gene pairs, but is evaluated based on the macro average F value and the micro average F value. The macro average F value is the calculated and averaged F value for each class, and the micro average F value is independent of different classes. Find the F value for all cases.

5 Conclusions

As we all know, in the power supply system, the voltage conversion and distribution of the system are achieved with the help of substations, which are the hubs connecting various power grids, organically linking different levels of power grids, and carrying out planned, step-by-step, purposeful control and diversion of electric energy, and its safety and stability directly affect the security of the entire power grid system. Therefore, investment in substations is very important and promotes the country's economic development. Therefore, it is necessary to use Granger causal calculation of the investment benefit of the substation.

This method has achieved higher estimation accuracy than existing methods in the comparative experiment of artificial data and actual data. In the existing model based methods that use regression models, the estimation accuracy varies greatly according to whether the regression model fits the data well, but the proposed method uses the same feature expression and the same classifier (in In the experiment, random forest), reached a sufficiently high estimation accuracy. Compared with the existing causal inference method RCC based on the classification of i.i.d. data, the proposed method also shows higher estimation accuracy. This result implies the validity of the proposed feature quantity representation.

In addition, in this article, we also describe a method for extending the proposed method so that Granger Causal relationship can be estimated from multivariate time series, and the effectiveness of the extended proposed method is verified through experiments using artificial data and actual data.

The future prospect is to deal with complex actual data. Specifically, it can handle non stationary time series. For nonstationary time series, it is necessary to consider 1) whether Granger Causal relationship and direction will not change, 2) whether Granger Causal relationship and direction change with time, so this case is not applicable. To cope with such complex problem settings, further expansion of the proposal method is a future issue.

In addition, in the results shown in Fig. 3 to Fig. 5, There is no guarantee that the feature vector will be obtained in this way. It is also very important as a future subject to study what kind of time series can be obtained based on the existence and direction of Granger Causal relationship from both theoretical and experimental aspects.

References

[1] V. Corradi, W. Distaso, A. Mele, Macroeconomic determinants of stock volatility and volatility premiums, Journal of Monetary Economics 60(2)(2013) 203-220.

[2] M.F. Ghalwash, Z. Obradovic, Early classification of multivariate temporal observations by extraction of interpretable shapelets, BMC bioinformatics 13(2012) 195.

[3] Y. Cao, Y. Li, S. Coleman, A. Belatreche, T.M. McGinnity, Adaptive Hidden Markov Model with Anomaly States for Price Manipulation Detection, IEEE transactions on neural networks and learning systems 26(2)(2015) 318-330.

[3] H. Chen, P. Tino, A. Rodan, X. Yao, Learning in the Model Space for Cognitive Fault Diagnosis, IEEE Transactions on Neural Networks and Learning Systems 25(1)(2014) 124-136.

[4] C. Hiemstra, J.D. Jones, Testing for Linear and Nonlinear Granger Causality in the Stock Price- Volume Relation, The Journal of Finance 49(5)(1994) 1639-1664.

[5] T. Górecki, M. Łuczak, Multivariate time series classification with parametric derivative dynamic time warping, Expert Systems with Application 42(5)(2015) 2305-2312.

[6] G.W. Schwert, Stock Returns and Real Activity: A Century of Evidence, The Journal of Finance 45(4)(1990) 1237-1257.

[7] H. Jaeger, M. Lukoševičius, D. Popovici, U. Siewert, Optimization and applications of echo state networks with leaky-integrator neurons, Neural Networks: The Official Journal of the International Neural Net-work Society & Japanese Neural Network Society 20(3)(2007) 335-352.

[8] J.P. Byrne, G. Fazio, N. Fiess, Primary commodity prices: Co-movements, common factors and fundamentals, Journal of Development Economics 101(2013) 16-26.
[9] J. Quevedo, H. Chen, M.À. Cugueró, P. Tino, V. Puig, D. García, R. Sarrate, X. Yao, Combining learning in model space fault diagnosis with data validation/reconstruction: Application to the Barcelona water network, Engineering Applications of Artificial Intelligence: The International Journal of Intelligent Real Time Automation 30(2014) 18-29.
[10]J.R. McCrorie, M.J. Chambers, Granger causality and the sampling of economic processes, Journal of Econometrics 132(2)(2006) 311-336.
[11]S. Ma, Y. Zheng, O. Wolfson, Real Time City-Scale Taxi Ridesharing, IEEE Transactions on Knowledge and Data Engineering 27(7)(2015) 1782-1795.
[12]M.F. Ghalwash, Z. Obradovic, Early classification of multivariate temporal observations by extraction of interpretable shapelets, BMC bioinformatics 13(2012) 195.
[13]N.M. Odhiambo, Trade Openness and Energy Consumption in Sub-Saharan African Countries: A Multi-variate Panel Granger Causality Test, Energy Reports 7(2021) 7082-7089.
[14] J.R. McCrorie, M.J. Chambers, Granger causality and the sampling of economic processes, Journal of Econometrics 132(2)(2006) 311-336.