2020 16th International Conference on
Control, Automation, Robotics and Vision (ICARCV)
December 13-15, 2020. Shenzhen, China

Sparse Causal Residual Neural Network for Linear and Nonlinear Concurrent Causal Inference and Root Cause Diagnosis

Jiawei Chen, Chunhui Zhao, *IEEE senior member*, and Youxian Sun

Abstract— **Reliable and effective fault diagnosis methods are necessary for complex industrial processes that consists of various units. After a process fault is detected, it remains a challenging task to locate the root cause unit and determine the propagation path of the fault. In this paper, a novel method, termed Sparse Causal Residual Neural Network (SCRNN), is proposed and applied for modern industrial root cause diagnosis. The advantage of SCRNN lies in that it can not only recognize linear and nonlinear causal relationships in parallel, but also automatically determine the causality lags and deduce the time delay of causal transmission. Besides, due to the specially designed sparse constraint and optimization algorithm, the SCRNN model can realize the function of key dependent variable selection, avoiding the high computational complexity and complicated procedure brought by pairwise comparison. The feasibility of the proposed method is illustrated through the benchmark TE process.**

I. INTRODUCTION

Corresponding to the rapid development of modern industrial technology, the complexity of the industrial system is also increasing. In order to ensure the safety of manufacturing and the stability of product quality, it is necessary to establish reliable and effective industrial fault diagnosis strategies [1]. Fault diagnosis aims to quickly detect process fault elements and determine the root causes of these faults by using appropriate models and inference methods [2]. Among these, the root cause diagnosis is a challenging task and has not yet been fully addressed.

In recent years, data-driven methods have shown unique advantages and popularity in the field of industrial fault diagnosis[3]-[7]. Among them, data-driven causal analysis methods, such as Granger Causality analysis[8], Bayesian networks[9]-[11], transfer entropy[12][13], aim to determine the propagation direction of faults by analyzing the causal relationship between process variables, thus deducing the key fault variables and root cause fault variables. Such methods often use statistical analysis to extract information from industrial data. Among them, the Granger Causality Test is a popular method derived from econometrics and has been applied to many other fields because of its strong interpretability. The G-causality Test is based on AR model, aiming to analyze the causality between two variables. However, the traditional G-causality analysis is only applicable to linear and stationary time series data. Besides, it determines the relationship between pairs of variables, which limits its usage scenarios. The nonlinear Granger causality based on radial basis function[14], GPR-based Granger Causality test [15] and transfer entropy(TE) can handle the problem of nonlinear. However, it still needs to compute the variables in pairs, and the procedure is complicated. Grouped Graphical Granger Modeling Methods[16] can excavate causal relationships among multiple variables based on the Vector Autoregression model and group lasso penalty. However, it's essentially a linear regression method. Recently, as an extension of VAR, Neural Granger Causality[17] was designed for multivariable nonlinear Granger analysis. However, for simple linear causal relations, such a deep learning method lacks interpretability and may mislead false results due to the risk of overfitting. Consider that modern industrial processes are often multivariable, in which linear and nonlinear causality coexist. These methods mentioned above have more or fewer limitations in application.

In this study, a novel data-driven strategy termed Sparse Causal Residual Neural Network (SCRNN) is proposed for industrial root cause fault diagnosis. Similar to the Vector Autoregression model, the SCRNN model takes the past information of the whole multiple time series as input, to predict the future information of a target variable through the regression method. Through the special network structure and sparse constraint methods, SCRNN can directly obtain the causal relationship between variables after optimization. The SCRNN structure is composed of two parts. The first part is a variable selection module. Its function is to perform a sparse linear transformation on input data and screen out key information of key variables. The second part is a fitting module, which is composed of a residual learning framework. Its function is to perform linear and nonlinear fitting and obtain the prediction result.

The main innovations and contributions of this work are summarized as below:

1) A novel causal analysis method is proposed, which can simplify the procedure of causality determination, and avoids the complicated procedure of pairwise comparison in multivariate cases.

2) The proposed SCRNN model can identify linear and nonlinear causality parallelly among complex multivariate data in modern industrial processes. Additionally, by designing the sparse constraint and optimization algorithm, the SCRNN model can determine the causality lag automatically.

Research supported by Zhejiang Key Research and Development Project (2019C01048), NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (No. U1709211), and Open fund of Science and Technology on Thermal Energy and Power Laboratory (No.TPL2019C03)

Jiawei Chen is with School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China.

Chunhui Zhao and Youxian Sun are with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China (Corresponding author: Chunhui Zhao, phone: +86-0571-87952301, e-mail: chhzhao@zju.edu.cn)

II. PRELIMINARY

A. Granger Causality Test

Granger causality originated in the field of econometrics, aiming to study whether a time series is causally affected by another time series based on predictability.

For two time series $X_1(t)$ and $X_2(t)$ from two stationary stochastic processes, $X_1(t) = (x_1(1), x_1(2), ..., x_1(n))$, $X_2(t) = (x_2(1), x_2(2), ..., x_2(n))$, a bivariate autoregressive (AR) model can be built as follows:

$$x_1(t) = \sum_{l=1}^{k} a_{11,l} x_1(t-l) + \sum_{l=1}^{k} a_{12,l} x_2(t-l) + e_1(t) \quad (1)$$

$$x_2(t) = \sum_{l=1}^{k} a_{21,l} x_1(t-l) + \sum_{l=1}^{k} a_{22,l} x_2(t-l) + e_2(t) \quad (2)$$

and a reduced model as follows:

$$x_1(t) = \sum_{l=1}^{k} b_{11,l} x_1(t-l) + e_{1(2)}(t) \quad (3)$$

$$x_2(t) = \sum_{l=1}^{k} b_{22,l} x_2(t-l) + e_{2(1)}(t) \quad (4)$$

where $a_{ij,l}, b_{ii,l}$ are the AR coefficients, and k is the model order which defines how many time lags to be included in the regression model. k can be selected by maximizing the Akaike information criterion (AIC)[18] or the Bayesian information criterion (BIC)[19].

$e_{i(j)}(t), e_i(t)$ are the prediction errors or residuals of the model. $e_{i(j)}(t)$ refers to the prediction error from a model that predicts the ith variable by excluding the jth variable. If the variance of $e_i(t)$ is significantly less than $e_{i(j)}(t)$, then the prediction accuracy of time series $X_i(t)$ is significantly improved when the previous X_j is included. So there is a causal effect from $X_j(t)$ to $X_i(t)$ .The difference of variances can be quantified by using the following ratio indicators:

$$F_{j \to i} = \ln\left(\frac{var(e_{i(j)})}{var(e_i)}\right) \quad (5)$$

When there is causal influence from $X_j(t)$ to $X_i(t)$, $F_{j \to i} > 0$, otherwise, $F_{j \to i} = 0$. Additionally, $F_{j \to i}$ can never be negative. The statistical significance of this effect can be tested by the F statistic.

Based on AR model, the traditional Granger causality test is applicable to the binary variable scenarios. In multivariable cases, Granger causality test is performed by a Vector Autoregression(VAR) model usually:

$$x_t = \sum_{l=1}^{k} \boldsymbol{A}^{(l)} x_{t-l} + e_t \quad (6)$$

where $\boldsymbol{A}^{(l)}$ is a $p \times p$ coefficient matrix that denotes how lag l affects the future evolution of the time series and e_t is a p-vector of error terms that satisfying $E(e_t) = 0$. In other words, e_t is mean zero noise. Time series j does not Granger cause time series i *iff* $\forall l, A_{ij}^{(l)} = 0$. Therefore, the focus of Granger causality analysis in a VAR model is to determine which values in $A^{(l)}$ are zeros over all lags. It can be figured out by solving a regression problem with lasso penalty.

As a conclusion, for multivariate linear and stationary time series, we can obtain the causal relationship between multivariate variables by solving a VAR problem.

III. METRODOLOGY

In this part, we begin with the idea of Non-linear Granger Causality and Neural Granger Causality method. Then we introduce the proposed SCRNN model in detail from its structure, function and application scenario.

A. Non-linear Granger Causality

The central assumption of Granger Causality is that causes precede their effects: if the addition of past information of time-series $\boldsymbol{X}$ can help to predict time series $\boldsymbol{Y}$, then $\boldsymbol{X}$ is the "Granger Cause" of $\boldsymbol{Y}$. Traditional Granger Test method is based on linear regression models and desigend for linear relations. Here, we introduce a more recent definition of Tank et al.[17] for non-linear Granger Causality:

Given a multivariate time series p-dimensional stationary time series $X \in \mathbb{R}^p$, $X = (x_1, ..., x_T)$ and a non-linear autoregression function g_j:

$$x_{tj} = g_j(x_{<t1}, ...x_{<ti}, ..., x_{<tp}) + e_{tj} \quad (7)$$

where $x_{<ti} = (..., x_{(t-2)i}, x_{(t-1)i})$ represents the past elements of series i and e_t represents the indenpent noise. Under this context, time-series i Granger causes j if g_j is not invariant to $x_{<ti}$, i.e. $if\ \exists x'_{<ti} \neq x_{<ti}$:

$g_j(x_{<t1}, ...x'_{<ti}, ..., x_{<tp}) \neq g_j(x_{<t1}, ...x_{<ti}, ..., x_{<tp})$

In other words, Nonlinear Granger Causality from series i to j means that the function g_j is dependent on $x_{<ti}$.

Such nonlinear regression function can be realized through multi-layer neural network or other methods. As an extention of VAR, Neural Granger Causality[17], on the basis of neural network framwork and sparse constraint, aims to extract the Granger causal structure in nonlinear multivariable scenarios.

Through the nonlinear multilayer perceptron(MLP), Neural Granger method uses the past information of multivariate time series to predict the future elements of a target variable:

$$h_t^1 = \sigma\left(\sum_{k=1}^{K} W^{1k} x_{t-k} + b^1\right) \quad (8)$$

$$x_{ti} = g_i(x_{<t}) + e_{ti} = w_O^T h_t^L + e_{ti} \quad (9)$$

The first neural network layer's weight reflects the causal dependence from input variables to the target variables, Therefore, by applying sparse constraints on this weights, the key dependent variables can be screened. Considere the regression accuracy and sprase constraint, the optimization objective fuction is as follws:

$$\min_{W}\left(\sum_{t=K}^{T}(x_{it}-g_i(x_{<t}))^2+\lambda\sum_{j=1}^{p}\|\boldsymbol{W}^{11}_{:j},...,\boldsymbol{W}^{\boldsymbol{IK}}_{:j}\|_2\right) \quad (10)$$

After solving this group lasso regression problem, we can get the first network layer's weight, which is a sparse matrix. The corresponding zero weight of a variable means that the information of this variable does not participate in the regression model. Namely, this variable is the Granger non-causality of the target variable. Therefore, the Neural Granger method learns nonlinear causal relationship by detecting the first neural network layer's weight is zero or not.

B. The Proposed Sparse Causal Residual Neural Network

This section introduces the proposed SCRNN model in detail, which is a neural network and VAR based method aims to determine the causal relationships and causality lags between multiple variables. As shown in fig. 1, The SCRNN model is composed of two parts. The first part is a variable selection module. Its function is to perform a sparse linear transformation on input data and screen out key information of key variables. The second part is a fitting module, which is composed of a residual learning framework [20]. Its function is to perform linear and nonlinear fitting and obtain the prediction result.

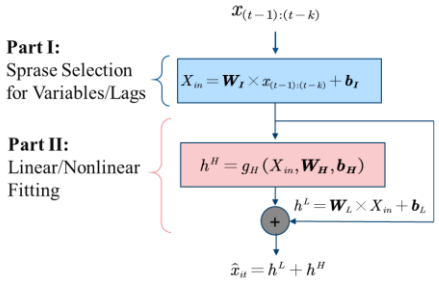

Fig. 1: Schematic of Sparse Causal Residual Neural Network

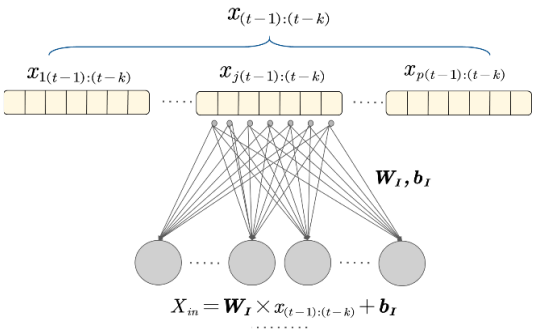

Fig. 2: Detailed Structure of Input Layer in SCRNN

The detailed structure of the input layer in SCRNN is shown in fig. 2. Given an original input data $x_{(t-1):(t-K)}$, the input layer selects both variables and their lags that help to predict by a sprase linear transformation :

$$X_{in}=\boldsymbol{W_I}\times x_{(t-1):(t-k)}+\boldsymbol{b_I} \quad (11)$$

Then there are two ways to transform selected data from input to ouput. In the nonlinear way, the transformation is realized by a multi-layer neural network as follows:

$$h^{h1}=\sigma(X_{in}) \quad (12)$$

$$h^H=\sigma(\boldsymbol{W}^H h^{h1}+\boldsymbol{b}^H) \quad (13)$$

where $\sigma(\cdot)$ is a nonlinear activation function such as $\mathrm{Re}Lu(x)=\max(0,x)$.

In the linear transformation way, the output is as follows:

$$h^L=\boldsymbol{W}^L X+\boldsymbol{b}^L \quad (14)$$

The final output of the network is a simple addition of linear transformation and nonlinear transformation output :

$$\hat{x}_{it}=h^H+h^L \quad (15)$$

1) Hierarchical Group Lasso Penalty for both Variable and Lag Selection

As shown in fig. 2, the structure of SCRNN is divided into two parts. The first part contains only the input layer, whose function is to select variables and lags sparsely by a linear transformation. So the group lasso penalty is applied to the first layer's weight parameter $\boldsymbol{W_I}$ to make sparse selection of variables and lags. The form of the hierarchical group lasso penalty term[21] is as follows:

$$\lambda\sum_{j=1}^{p}\sum_{l=1}^{k}\|\boldsymbol{W}^{\boldsymbol{I}l}_{:j},...,\boldsymbol{W}^{\boldsymbol{I}k}_{:j}\|_2 \quad (16)$$

In combination with Proximal Gradient Descent(PGD) optimization algorithm[22], the hierarchical penalty leads to solutions such that for each candidate dependent variables variable x_j:

$$\exists\, l'\in\{0,1,2,...,k\},\ for\ \forall\ 0<l\leqslant l'\ ,\boldsymbol{W}^{\boldsymbol{I}l}_{:j}\neq 0;$$
$$\forall\ l'<l\leqslant k\ ,\boldsymbol{W}^{\boldsymbol{I}l}_{:j}=0$$

Thus, with the help of the hierarchical group lasso penalty, SCRNN can effectively select the lag of each causal relationship. Consider the most extreme case that $l'=0$, which means that $\boldsymbol{W}^{\boldsymbol{I}l}_{:j}$ is equal to zero across all l. It turns out that x_j is not the key dependent variable of x_i. Therefore such a penalty term can also select for key dependent variables. In practice, it is generally necessary to set k to a large value to ensure that no causal relationship is missed.

2) The Relationship between Linear Shortcut Layer and VAR

Since causal inference focuses more on the relationship between variables than on the fitting accuracy of regression. Therefore, if only the nonlinear models are used to fit the linear relationship, it may have better fitting accuracy but lead to incorrect causal inference results. In the SCRNN model, the linear shortcut layer is designed to avoid this problem by capture the linear coupling between variables like the VAR model.

For the input data $x_{(t-1):(t-K)}$, after transformation by the input layer and shortcut layer, the output is as follows:

$$\begin{aligned}h^L&=\boldsymbol{W_L}(\boldsymbol{W_I}\times x_{(t-1):(t-k)}+\boldsymbol{b_I})+\boldsymbol{b_L}\\&=(\boldsymbol{W_L}\boldsymbol{W_I})\times x_{(t-1):(t-k)}+C\end{aligned} \quad (17)$$

where C is a constant term that doesn't depend on $x_{(t-1):(t-k)}$, $\boldsymbol{W_I}$ and $\boldsymbol{W_L}$ are trainable variables that can be updated during the iteration of optimization. The product of $\boldsymbol{W_I}$ and $\boldsymbol{W_L}$ can be viewed as a new matrix. Therefore, it can be proven that the combination of the input layer and the

linear shortcut layer is also a simple linear transformation, just like VAR. That is, if the causal relationship between variables is simply linear, it can be identified by the linear shortcut layer of SCRNN easily.

As a conclusion, the design of the shortcut layer reduces the risk of false causal inference results when simple linear relationships are possibly overfitting by the multi-layer neural networks.

3) Optimization objective functions for SCRNN

The objective optimization function of SCRNN consists of two parts. Like the VAR model, SCRNN needs to minimize prediction residuals and to ensure the fitting accuracy. Besides, it's needful to minimize the hierarchical group lasso penalty term to ensure an interpretable sparse solution. Considering the above two points, the objective optimization function is as follows:

$$\min\left(\sum_{t=k}^{T}(x_{it}-\hat{x}_{it})^2+\lambda\sum_{j=1}^{p}\sum_{l=1}^{k}\|\boldsymbol{W}^{\boldsymbol{I}^l}_{:j},...,\boldsymbol{W}^{\boldsymbol{I}^k}_{:j}\|_2\right) \quad (18)$$

A. SCRNN for Causal Inference of variables and lags

This section introduces the rule of causal inference based on SCRNN model. After optimization, we get the Input Layer's weight parameters $\boldsymbol{W_I}$. By observing the element values of the matrix $\boldsymbol{W_I}$, the causality among the variables and the lag of each causal relationship can be determined.

1) Method for determining key dependent variables of each single variable

Given that the model's input is a time series of length k: $x_{(t-1):(t-k)}=\{x_{1(t-1):(t-k)},...,x_{j(t-1):(t-k)},...,x_{p(t-1):(t-k)}\}$, and output is $\hat{x}_{it}$, the prediction of x_{it}.

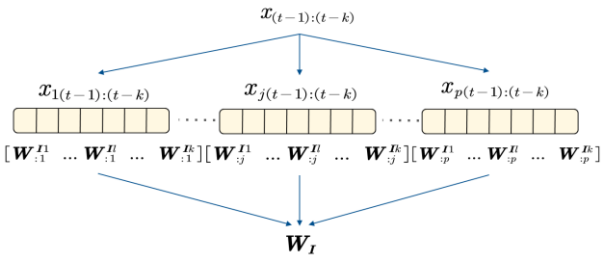

Fig. 3: The Corresponding Relationship between Input Layer's weight matrix and Input Data

Because of the sparse constraint added to $\boldsymbol{W_I}$, $\boldsymbol{W_I}$ will be a sparse matrix after optimization, and it can be written as a partitioned matrix. Fig. 3 shows the corresponding relationship betwween $\boldsymbol{W_I}$'s partitioned form and input data $x_{(t-1):(t-k)}$. The relationship and element value of $\boldsymbol{W_I}$ will reflect whether the input variable can help predict our output value, so it can screen out the key dependent variables.

We can derive the rule for determining the causal relationship between variables as follows:

- $if\ \forall l\in\{1,2,...,k\},\ \boldsymbol{W}^{\boldsymbol{I}^l}_{:j}=0,$

then variable x_j does not the Granger cause of x_i.

Specifically speaking, if $\boldsymbol{W}^{\boldsymbol{I}^l}_{:j}$ is zero for all $l<k$, then the information of x_j does not participate in the regression process of variable x_i. That means the addition of x_j cannot improve the prediction accuracy of x_i. This is analogous to the Granger idea of causality test, which illustrates that variable x_j does not the Granger cause of variable x_i.

2) Method for determining the lag of each causal relationship

Like the method to determine the key dependent variable, we can also determine the delay of each causal relationship by checking $\boldsymbol{W_I}$. For a variable x_j and its corresponding Input Weight $[\boldsymbol{W}^{\boldsymbol{I}^1}_{:j}\ ...\ \boldsymbol{W}^{\boldsymbol{I}^l}_{:j}\ ...\ \boldsymbol{W}^{\boldsymbol{I}^k}_{:j}]$:

- $\exists l'\in\{1,2,...,k\},\ if\ \forall\ 0<l\leqslant l'\ ,\ \boldsymbol{W}^{\boldsymbol{I}^l}_{:j}\neq 0,$
  $and\ \forall\ l'\ <l\leqslant k\ ,\ \boldsymbol{W}^{\boldsymbol{I}^l}_{:j}=0$

That the lag of causal relationship $x_j\rightarrow x_i$ is l'.

B. Online Fault Diagnosis Prosecure based on SCRNN

In this section, an online fault diagnosis procedure is introduced. The flow chart of fault diagnosis is shown in fig.4 . In practical industrial applications, through various monitoring methods, the faults in the industrial process can be detected and the fault variables can be determined meanwhile. After determining the fault variables, we can find out the potential root causes according to the proposed SCRNN model, and then conduct the isolation and recovery of the faults.

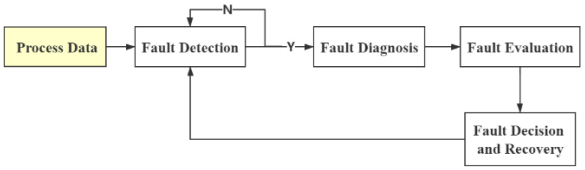

Fig. 4: Flowchart of fault diagnosis

The detailed root cause diagnosis procedure based on SCRNN is as follows:

Setp1: Take the multivariate time series of fault variables as the input data ***X***, and carry out a Maximum and minimum normalization for the input data ***X***.

Step2: Orderly select one dimension of the input data as the output target variable. Then use SCRNN model to obtain the key dependent variables and causality lags of every target variable. Given that the number of the input fault variables is *p*, so there are *p* models needed to be established.

Step3: After identifying the key dependent variables of every single variable, the global causal relationship among all variables can be obtained. Then the causal transmission route and the root cause can be determined.

In the industrial process, the root cause fault variable is often the first unit to fail, and the fault propagates to other units along with the material and signal flows. Therefore, they are usually upstream in the propagation path of a fault.

To sum up, given the multivariate time series of fault variables as input data, the SCRNA-based diagnosis procedure can output the potential root causes of the fault.

IV. Case Analysis

A. Tennessee Eastman (TE) Process

The TE Process[23] is a benchmark for evaluating the effectiveness of various fault detection and diagnosis methods. It has been widely used in the field of process monitoring and fault diagnosis. The TE process is mainly composed of a reactor, condenser, stripper, vapor-liquid separation tower, compressor, and other operating units. There are four reactants in the TE process: A, D, E, and C, all of which contain A small amount of inert gas B. There are 41 measured variables(include 22 continuous process variables and 19 component variables) and 12 manipulated variables recorded for data collection with a sampling interval of 3 min. In the TE process, the real situation of fault occurrence and normal operation is simulated by pre-setting different operation working states, including normal operation state and 21 different fault states. More detail about TE process can be found in the literature [23].

In this paper, we mainly focus on the first case IDV(1), This is a fault caused by the change of A/C feed ratio in stream 4, which results in the reduction of component A in the recycling stream. The composition controller effectively reacts and increases the flow rate (x_1) of the A feed in the stream1 by adjusting the corresponding valve opening x_{44}, which raises the level of the reactor, changes the material residence time, and affects many other process variables. The flow rate of flow 4 (x_4) is affected by the level controller, which further changes the stripper pressure (x_{16}) and temperature (x_{18}). Then, the product separator pressure (x_{13}) and reactor pressure (x_7) are affected. At the same time, due to the control of the stripping tower temperature, the stripping tower steam valve (x_{50}) is adjusted to control the stripping tower steam flow (x_{19}). The trajectories of several selected variables are shown in Fig. 5.

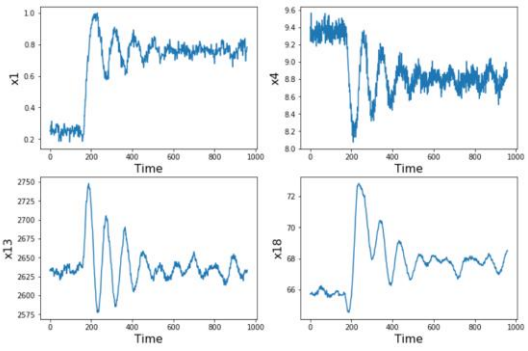

Fig. 5: Variable trajectories in TE process IDV(1) scenario(x_1, x_4, x_{13}, x_{18})

From the above analysis, it is clear that x1 and x44 are the variables closest to the root cause of this fault. Additional faulty variables are located downstream of the fault propagation paths. In IDV(1), the candidate set of faulty variables and their description are as a table following:

TABLE I IDV(1) Main Fault Variables and Their Deicription

ID	Variable Description	Variable Type
x_1	A feed (stream 1)	Measure
x_4	A and C feed (stream 4)	Measure
x_7	Reactor pressure	Measure
x_{13}	Product separator pressure	Measure
x_{16}	Stripper pressure	Measure
x_{18}	Stripper temperature	Measure
x_{19}	Stripper steam flow	Measure
x_{44}	MV to A feed flow (stream 1)	Manipulation
x_{50}	Stripper steam value	Manipulation

1) Determine key dependent variables and causal lag for a single variable in TE process IDV(1)

In this part, we introduce the specific procedure of determining the key dependent variable and the causal relationship lag for a selected target variable.

As an example, we take x_{13} (a measured variable of stripper pressure) as the target variable. Set the training lag *k* as 6, the key dependent variables and causality lag of x_{13} in IDV(1) based on SCRNN model are shown Fig.8 . A blue block in row *i* represents the variable corresponding row *i* is the key dependent variable of x_{13}. In other words, the corresponding variable is the Granger cause of x_{13}. Besides, the blue block can also indicate the causality lag of each causal relationship.

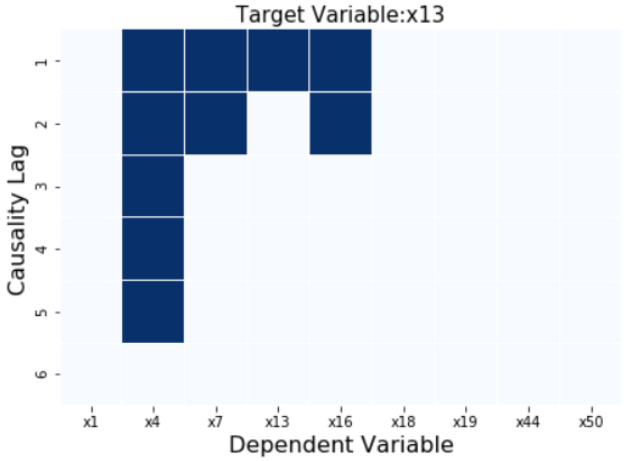

Fig.6: Key dependent variables and Causality lag of x_{13}

As indicated by Fig.6 , for target variable x_{13}, there are three dependent variables: x_4, x_7, x_{16}. It means that in the propagation path of the fault they are upstream variables compared with x_{13}. Besides, the causality lag of each relation differs. Causality lag based on variable x_4 to x_{13} is the maximum, which means variable x_4 have the most profound impact on x_{13}.

1) Causal relationship among all main variables in TE process IDV(1)

In this part, we analyze the causal relationship among all main variables in IDV(1) by searching and analyzing the key dependent variables of each single target variable. Fig 7 shows the global causality of IDV(1) based on SCRNN method, the horizontal axis represents the key dependent variables of each target variable. A blue block in row *i* column *j* represents variable *j* is the Granger cause of variable *i*. It should be noted that when considering the causal relationship between variables, it's not necessary to take the target variable itself into consideration, so the diagonal elements in the variable relation matrix are set to zero.

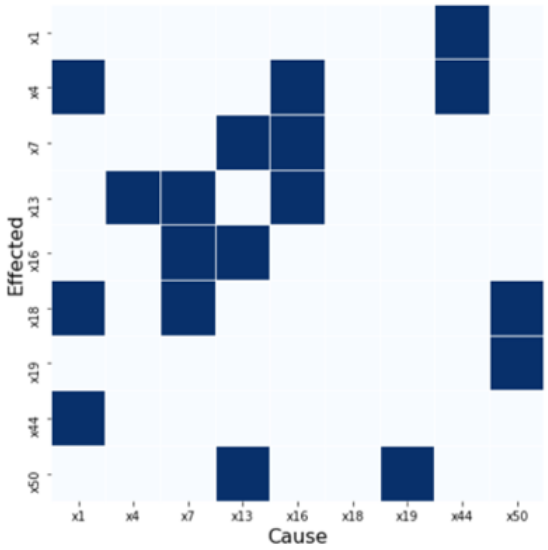

Fig.7: Casual graph of TE Process IDV(1) based on SCRNN

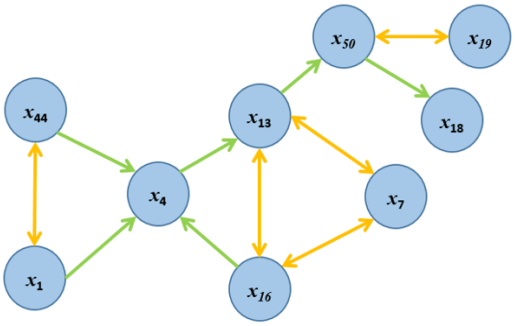

Fig.8: Casual graph of TE Process IDV(1) based on SCRNN

Through the causal graph shown in Fig7, we can draw the causal transmission path in Fig8, which also shows the propagation path of the fault between variables, from the upstream root cause to the downstream. In fig 10, the single arrow indicates the causal direction from cause to effect, while the double-head arrow represents the reciprocal causal relation. As clearly indicated by Fig.8 , x_1, x_{44} and x_4 were located upstream of the fault propagation path. x_1 and x_{44} are diagnosed as the root cause. There is a reciprocal causal relation between x_1 and x_{44}, and a causal direction from x_1 to x_4. The fault propagation path obtained by SCRNN model basically conforms to the mechanism description of the fault correctly.

V. Conclusion

In this paper, an SCRNN model is proposed for root cause fault diagnosis. Due to the specially designed sparse constraint and neural network structure, the model can automatically determine key causal relationships between variables and causality lags of these relationships. SCRNN based root cause diagnosis method avoids the complicated procedure caused by pairwise comparison, and can well adapt to the multivariable cases in modern industrial processes. Experimental results illustrate the superiority of this method.

VI. References

[1] C. Zhao and F. Gao, "Critical-to-Fault-Degradation Variable Analysis and Direction Extraction for Online Fault Prognostic", IEEE Transactions on Control Systems Technology, vol. 25, no. 3. IEEE Transactions on Control Systems Technology, pp. 842–854, 2017.

[2] R. Isermann, Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems. Springer Science & Business Media, 2011.

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

[4] S. Zhao, B. Huang, and F. Liu, "Fault Detection and Diagnosis of Multiple-Model Systems With Mismodeled Transition Probabilities", IEEE Transactions on Industrial Electronics, vol. 62, no. 8. IEEE Transactions on Industrial Electronics, pp. 5063–5071, 2015.

[5] W. Yu and C. Zhao, "Online Fault Diagnosis in Industrial Processes Using Multimodel Exponential Discriminant Analysis Algorithm", IEEE Transactions on Control Systems Technology, vol. 27, no. 3. IEEE Transactions on Control Systems Technology, pp. 1317–1325, 2019.

[6] Z. Chai and C. Zhao, "A Fine-Grained Adversarial Network Method for Cross-Domain Industrial Fault Diagnosis", IEEE Transactions on Automation Science and Engineering, vol. 17, no. 3. IEEE Transactions on Automation Science and Engineering, pp. 1432–1442, 2020.

[7] L. Feng and C. Zhao, "Fault Description Based Attribute Transfer for Zero-Sample Industrial Fault Diagnosis", IEEE Transactions on Industrial Informatics. IEEE Transactions on Industrial Informatics, pp. 1–1, 2020.

[8] C. Granger, "Investigating Causal Relationships by Econometric Models and Cross-Spectral Methods', Econom trica," ed: July, 1969.

[9] R. G. Cowell, P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, Probabilistic networks and expert systems: Exact computational methods for Bayesian networks. Springer Science & Business Media, 2006.

[10] G. Weidl, A. L. Madsen, S. J. C. Israelson, and c. engineering, "Applications of object-oriented Bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes," vol. 29, no. 9, pp. 1996-2009, 2005.

[11] W. Yu and C. Zhao, "Online Fault Diagnosis for Industrial Processes With Bayesian Network-Based Probabilistic Ensemble Learning Strategy", IEEE Transactions on Automation Science and Engineering, vol. 16, no. 4. IEEE Transactions on Automation Science and Engineering, pp. 1922–1932, 2019.

[12] T. J. P. r. l. Schreiber, "Measuring information transfer," vol. 85, no. 2, p. 461, 2000.

[13] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. J. I. t. o. c. s. t. Thornhill, "Finding the direction of disturbance propagation in a chemical process using transfer entropy," vol. 15, no. 1, pp. 12-21, 2006.

[14] N. Ancona, D. Marinazzo, and S. Stramaglia, "Radial basis function approach to nonlinear Granger causality of time series", Physical Review E, vol. 70, no. 5. Physical Review E, 2004.

[15] H.-S. Chen, C. Zhao, Z. Yan, and Y. Yao, "Root Cause Diagnosis of Oscillation-Type Plant Faults Using Nonlinear Causality Analysis", IFAC-PapersOnLine, vol. 50, no. 1. IFAC-PapersOnLine, pp. 13898–13903, 2017.

[16] A. C. Lozano, N. Abe, Y. Liu, and S. Rosset, "Grouped graphical Granger modeling methods for temporal causal modeling", 2009.

[17] A. Tank, I. Covert, N. Foti, A. Shojaie, and E. J. a. p. a. Fox, "Neural granger causality for nonlinear time series," 2018.

[18] H. J. I. t. o. a. c. Akaike, "A new look at the statistical model identification," vol. 19, no. 6, pp. 716-723, 1974.

[19] G. J. T. a. o. s. Schwarz, "Estimating the dimension of a model," vol. 6, no. 2, pp. 461-464, 1978.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[21] W. B. Nicholson, I. Wilms, J. Bien, and D. S. J. a. p. a. Matteson, "High dimensional forecasting via interpretable vector autoregression," 2014.

[22] N. Parikh, S. J. F. Boyd, and T. i. optimization, "Proximal algorithms," vol. 1, no. 3, pp. 127-239, 2014.

[23] J. J. Downs, E. F. J. C. Vogel, and c. engineering, "A plant-wide industrial process control problem," vol. 17, no. 3, pp. 245-255, 1993.