IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DATA-DRIVEN MONITORING, FAULT DIAGNOSIS AND CONTROL OF CYBER-PHYSICAL SYSTEMS

Received December 12, 2017, accepted January 13, 2018, date of publication January 23, 2018, date of current version March 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2795535

Process Monitoring Based on Multivariate Causality Analysis and Probability Inference

XIAOLU CHEN, JING WANG, (Member, IEEE), AND JINGLIN ZHOU, (Member, IEEE)
College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

Corresponding author: Jing Wang (e-mail: jwang@mail.buct.edu.cn)

This paper was supported in part by the National Natural Science Foundation of China under Grant 61573050 and Grant 61473025, in part by the Open-Project Grant funded by the State Key Laboratory of Synthetical Automation for Process Industry, Northeastern University, under Grant PAL-N201702.

ABSTRACT System security is one of the key challenges of the cyber-physical systems. Bayesian approach can estimate and predict the potentially harmful factors of the general system, but it has many limitations that can lead to undesirable effects in the complex systems. This paper presents a new modeling and monitoring framework to avoid the traditional Bayesian network disadvantage. A multivariate causal analysis method is proposed to establish a compact system structure. Combined with network parameter learning, we constructed a corresponding multivariate alarm predict graph model, in which the qualitative and quantitative relationships among the process variables are revealed distinctly. Then this model is used to accurately predict the future possible alarm events via the probability inference. Similarly, it also can be used to detect faults and find the source of the fault. The effectiveness of the proposed method is verified in public data sets and the Tenessee Eastman process. Simulation results show that the established causal relationship is completely consistent with the actual mechanism, and the alarm state of the critical variable is accurately predicted.

INDEX TERMS Alarm prediction, multivariate causality analysis, process monitoring modeling, parameter learning.

I. INTRODUCTION

Modern industrial process is a multidimensional complex system of integrated computing, network and physical environments, namely Cyber-Physical Systems (CPS) [1]. It realizes the real-time sensing, dynamic control and information service of large-scale engineering system through the deep collaboration of 3C (Computer, Communication and Control) technology [2], [3]. It is hard to build its precise mechanism (or mathematical) model for system monitoring. Fortunately, the development of information technology has provided new ideas to solve this problem. A large amount of process variables are measured regularly to promote data-driven statistical process monitoring techniques [4]–[7]. These methods extract the useful information from the acquired data and build a statistical monitoring model to describe the operating status of the entire process. Many excellent results of fault detection and diagnosis have been obtained based on multivariate statistical analysis [8]–[11]. However, the traditional multivariate statistical monitoring methods, such as principal component analysis (PCA) [12], independent component analysis (ICA) [13], [14], Fisher discriminant analysis (FDA) [15] and partial least-squares (PLS) [16], rely too heavily on the process data to focus on the intrinsic characteristics and connections among the system variables. How to figure out the interrelationships within system? Can we simplify the description of the complex process based on the variable interrelationships and process data simultaneously? To answer these questions, we devote to combine the effects of physical relationships and process data to make the monitoring model more persuasive.

As we all known, there are many ways to describe the system characteristic according to the observational data and expert knowledge, such as graph model [17], neural network model [18], fuzzy model [19]. The graph model is composed of points and lines to describe the system structure and the causal relationships among variables. It provides an effective method for studying various systems, especially the complex systems. Bayesian network(BN), a typical graph model, is the main method to deal with the knowledge representations and uncertainties based on the probability theory [20]. BN obtains the causality and probability within the process components and the system variables from the prior knowledge and

process data. The Bayesian network consists of the structure learning and the parameter learning, in which the structure learning aims to determine the causalities within system variables and the parameter learning will reveal the quantitative relationship of these causalities. BN has been successfully applied to fault diagnosis, financial analysis, automatic target recognition, military and many other areas [21], [22].

Bayesian network structure learning has some deficiencies when it is applied to the CPS system, such as complex training mechanism and variable causalities. In order to simplify the network structure, lots of assumptions should be presupposed and it inevitably causes the loss of generality. A variety of causal discovery methods have been proposed in recent years that were claimed to be able to find the causalities [23]–[26]. Usually, a generative model (linear or nonlinear) is built to explain the data generating process, i.e. the causalities. The most typical is the linear non-gaussian acyclic model (LiNGAM) proposed by Shimizu [27]. Its full structure was shown to be identifiable without pre-specifying a causal order of the variables. The improve LiNGAM method can estimate a causal order of variables without any prior structure knowledge, and provide better statistical performance [28]. A nonlinear causality of a pair of variables is discovered in [29], [30], and the proposed method shows a limitation when dealing with the multivariate variables. All these approaches exploit the complexity of the marginal and conditional probability distributions in one way or the other. Despite the large number of methods for bivariate causal discovery has been proposed over the last few years, their practical performance has not been studied very systematically. Most CPSs do not meet the linear and bivariate assumptions, the effectiveness of these methods have yet to be improved.

Alarm prediction and fault backtracking play an important role in the safety monitoring of the industrial CPSs. The research on the alarm system mainly focuses on the optimization of alarm structure and alarm threshold [31], [32], and few focus on the alarm prediction problem. It is very valuable to establish an alarm predict model and predict the alarm values of some key variables based on the internal variables relationship. Bayesian network model and the relative inference methods will show significant advantage in the alarm prediction and fault backtracking [33], [34]. There are several bottlenecks in CPSs' safety monitoring, including the discovery of multivariate causalities, simplification of internal structure and precise prediction of fault variables.

We propose a more generalized multivariate post-nonlinear acyclic causal model for the complex industrial process in this paper. The proposed multivariate post-nonlinear acyclic causal model, as an alternative network, can easily find the multi-variable's causality. It shows more compact structure and consistency with mechanism, compared to the Bayesian network structure. In addition, it avoids the complex learning mechanism of traditional Bayesian networks, thus making it easier to implement without compromising accuracy. The complete monitoring model is established via combining the causal discovery algorithm and the Bayesian parameters learning algorithm. The qualitative and quantitative relationships among the process variables are revealed to the greatest extent. Then this model is used to accurately predict the operation status of the critical process variables. Similarly, it also can be used to detect faults and find the source of the faults. The overall block diagram of our work is shown in Figure 1.

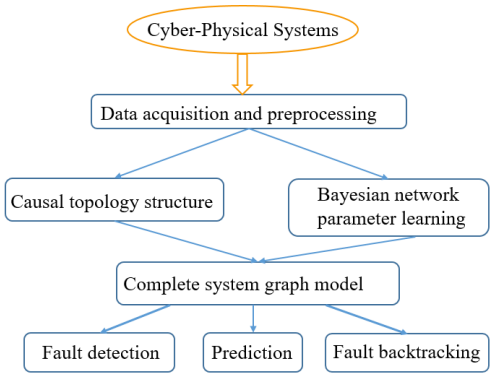


FIGURE 1. Overall design block diagram.

The paper is structured as follows. Section 2 is the methodology and the following problems are discussed in detail: how to find the causal relationship for the nonlinear continuous multiple variables; Bayesian network parameter learning and probabilistic inference for the proposed multivariate post-nonlinear acyclic causal model. The verification of the proposed method is shown in Section 3, and two experiments are finished aiming at the public data sets and the TE simulate platform. Finally, Section 4 draws the conclusions.

II. METHODOLOGY

A. ESTABLISHING CAUSAL STRUCTURE

Model-based causal discovery assumes a generative model to explain the data generating process. When the existing knowledge about the data model is unavailable, the assumed model should be sufficiently general so that it can be adapted to approximate the real data generation process. Furthermore, the model should be identifiable such that it could distinguish causes from effects. A nonlinear and multivariable process always possesses the following three characteristics:

1. The multivariate causalities are usually nonlinear.
2. The final target variable is affected by its cause variables and some noise who is independent from the causes.
3. Sensors or measurements may introduce nonlinear distortions into the observed value of the variables.

To discover the causality of multivariable in CPSs, a more generalized multivariate post-nonlinear acyclic causal model with inner additive noise is proposed. The model is in the form of graph theory and Bayesian network structure. Assume that there is a directed acyclic graph (DAG) to represent the relationship among multiple observed variables. Mathematically,

the generating process of X_i is

$$X_i = f_{i,2}(f_{i,1}(PA_i) + e_i), \tag{1}$$

where the observed variables X_i,$i = \{1, 2, \ldots, n\}$ is arranged in a causal order, such that no later variable causes any earlier variable. PA_i is the direct causes of X_i. $f_{i,1}$ denotes the nonlinear effect of the causes, and $f_{i,2}$ denotes invertible post-nonlinear distortion in variable X_i. e_i is the independent disturbance which is a continuous-valued random variable with non-gaussian distributions of non-zero variances. Model (1) satisfies the aforementioned three conditions: function $f_{i,1}$ accounts for the nonlinear effect of the causes PA_i; e_i is the noise effect during the transmission from PA_i to X_i; invertible function $f_{i,2}$ refects the nonlinear distortion caused by the sensor or measurement.

Randomly select a pair of variables X_i and X_j, $i,j = \{1, 2, \ldots, n\}$ from a multivariable system, respectively. Assume that the pair (X_i, X_j) has the causal relation $X_i \to X_j$. It's data generating process can be described in a generated model,

$$X_j = f_{j,2}(f_{j,1}(X_i) + e_j), \tag{2}$$

where e_j is independent from X_i. Define $s_i \triangleq f_{j,1}(X_i)$, $s_j \triangleq e_j$, and s_i is independent from s_j.

Rewrite the generating process $X_i \to X_j$ as follows:

$$\begin{aligned} X_i &= f_{j,1}^{-1}(s_i), \\ X_j &= f_{j,2}(s_i + s_j). \end{aligned} \tag{3}$$

X_i and X_j in (3) are post-nonlinear (PNL) mixtures of independent sources s_i and s_j. So the PNL mixing model can be seen as a special case of the general nonlinear independent component analysis (ICA) model. Here we use non-linear ICA method to solve this problem (3) [35].

Generally there are two possibility to describe the causal relation between any two random variables X_i and X_j, ($X_i \to X_j$ and $X_j \to X_i$). We should identify the correct relation by judging which one satisfies the assumed model (2). If the causal relation is $X_i \to X_j$ (i.e., X_i and X_j satisfy the model (2)), we can invert the data generating process (2) to recover the disturbance e_j, which is expected to be independent from X_i. Two steps are used to examine possible causal relationships between variables.

In the first step, recover the disturbance e_j corresponding to the assumed causal relation $X_i \to X_j$ based on the constrained nonlinear ICA. If this causal relation holds, there exist nonlinear functions $f_{j,2}^{-1}$ and $f_{j,1}$, we have

$$e_j = f_{j,2}^{-1}(X_j) - f_{j,1}(X_i), \tag{4}$$

where e_j is independent from X_i. Thus perform nonlinear ICA using the structure in Figure 2 and the outputs of system (5) are

$$\begin{aligned} Y_i &= X_i, \\ Y_j &= e_j = g_j(X_j) - g_i(X_i)). \end{aligned} \tag{5}$$

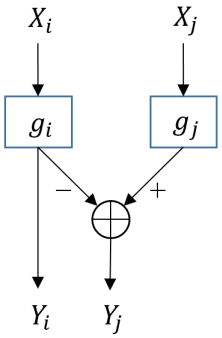


FIGURE 2. The constrained nonlinear ICA system used to verify if the causal relation $x_i \to x_j$ holds.

The nonlinearities g_i and g_j is modeled by Multi-layer perceptrons (MLP's), and the parameters in g_i and g_j are learned by making Y_i and Y_j as independent as possible, i.e., minimizing the mutual information between Y_i and Y_j,

$$I(Y_i, Y_j) = H(Y_i) + H(Y_j) - H(Y), \tag{6}$$

where $H(Y)$ is the joint entropy of $Y = (Y_i, Y_j)^T$,

$$\begin{aligned} H(Y) &= -Elogp_Y(Y) \\ &= -Elogp_Y(X) - log\,|J| \\ &= H(X) + Elog\,|J|. \end{aligned} \tag{7}$$

The joint density of $Y = (Y_i, Y_j)^T$ is $p_Y(Y) = p_X(X)/|J|$. J is the Jacobian matrix of the transformation from (X_i, X_j) to (Y_i, Y_j), i.e.,

$$\begin{aligned} J &= \frac{\partial(Y_i, Y_j)}{\partial(X_i, X_j)}, \\ |J| &= \begin{vmatrix} 1 & 0 \\ g_i' & g_j' \end{vmatrix} = |g_j'|. \end{aligned} \tag{8}$$

Substitute (8) and (9) into (7), we have

$$\begin{aligned} I(Y_i, Y_j) &= H(Y_i) + H(Y_j) - Elog|J| - H(X) \\ &= -Elogp_{Y_i}(Y_i) - Elogp_{Y_j}(Y_j) - Elog|g_j'| - H(X), \end{aligned} \tag{9}$$

where $H(X)$ does not depend on the parameters in g_i and g_j and can be considered as constant. The minimization problem (10) is solved by gradient-descent methods, and the details of the optimization are skipped.

In the second step,verify if the estimated disturbance Y_j is independent from the assume cause Y_i based on the statistical independence tests. We adopt the kernel-based statistical test [36], with the significance level $= 0.01$. Denote the test statistics as $test_{i\to j}$. If $test_{i\to j} > test_{j\to i}$, indicating that Y_i and Y_j are not independent, that is $X_i \to X_j$ does not hold. Repeat the above procedure (with X_i and X_j exchanged) to verify if $X_j \to X_i$ holds. If $test_{i\to j} < test_{j\to i}$, usually we can conclude that X_i causes X_j. g_i and g_j provide an estimate of $f_{j,1}$ and $f_{j,2}^{-1}$, respectively.

For a complex CPSs, there are n process variables. Following a test sequence, $X_1 \rightarrow X_2, X_1 \rightarrow X_3, \ldots, X_{n-1} \rightarrow X_n$, we need to test the N group statistics.

$$N = n + (n-1) + (n-2) + \ldots + 1 = \frac{n(n-1)}{2}. \quad (10)$$

The total computation is in direct proportion to $2 \times N$. In the simulation, in order to facilitate the display, we select the eight variables to calculate. As the number of variables increases, the amount of computation will increase as well. The measured statistics in the positive order (or in the reverse order) are stored as

$$\begin{aligned} A &= [test_{X_1 \rightarrow X_2}, test_{X_1 \rightarrow X_3}, \ldots, test_{X_{n-1} \rightarrow X_n}], \\ B &= [test_{X_2 \rightarrow X_1}, test_{X_3 \rightarrow X_1}, \ldots, test_{X_n \rightarrow X_{n-1}}]. \end{aligned} \quad (11)$$

Comparing the corresponding elements of the vectors A and B, the causal direction of this pair of variables is determined by finding a smaller statistic. Finally, we can find the causality of all variables by using cyclic search, and integrate it into a DAG.

B. PARAMETER LEARNING OF CAUSALITY BAYESIAN NETWORK

The multivariate causality model gives a framework similar to Bayesian networks to find the internal structure of the complex systems. This graphical structure expresses the causal interactions and direct/indirect relations as probabilistic networks. Its parameter represents the intensity of the complex inter-relationships among the cause-effect variables.

Consider a finite set $U = \{X_1, \ldots, X_n\}$ of discrete random variables where each variable X_i may take on several discrete status from a finite set. A Bayesian network is an annotated directed acyclic graph that encodes a joint probability distribution over a set of random variables U. Formally, a Bayesian network for U is a pair $B = \langle G, \Theta \rangle$. G is a directed acyclic graph whose vertices is correspond to the random variables $X_1, \ldots, X_n$. Θ is the parameters set that quantifies the network with $\theta_{ijk} = p(x_i^k | pa_i^j)$ and $\sum_k \theta_{ijk} = 1$, where x_i^k is the discrete status of X_i and pa_i^j is one of components in the complete parent set PA_i of X_i in G. Every variable X_i is conditionally independent of its non-descendants given its parents (Markov condition). The joint probability distribution over set U is

$$P_B(X_1, \ldots, X_n) = \prod_{i=1}^{n} P_B(X_i | PA_i) = \prod_{i=1}^{n} \theta_{X_i | \prod pa_i}. \quad (12)$$

The parameters of the causality Bayesian network are mainly learned from the sample data statistics analysis. The maximum likelihood estimation method (MLE) is one of the most classical and effective algorithms in parameter learning.

Given a dataset $D = \{D_1, \ldots, D_N\}$ of all BN nodes, the goal of parameter learning is to find the most probable values for Θ. These values best explain the dataset D, which can be quantified by the log likelihood function $logp(D|\theta)$, denoted $L_D(\theta)$. Assume that all samples are drawn independently from the underlying distribution. According to the conditional independence assumptions of BNs, we have

$$L_D(\theta) = log \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{n_{ijk}}, \quad (13)$$

where qi is the number of combinations of the parent nodes pa_i^j, and r_i is the number of the node X_i status. n_{ijk} indicates how many elements of D contain both x_i^k and pa_i^j. If the dataset D is complete, MLE method can be described as a constrained optimization problem:

$$\begin{aligned} &max\ L_D(\theta), \\ &st.\ g_{ij}(\theta) = \sum_{k=1}^{r_i} \theta_{ijk} - 1 = 0,\ \forall\, i = 1, \ldots, n, \\ &\forall\, j = 1, \ldots, q_i. \end{aligned} \quad (14)$$

The global optimum solution is

$$\theta_{ijk} = \frac{n_{ijk}}{n_{ij}}, \quad (15)$$

where $n_{ij} = \sum_{k=1,\ldots,r_i} n_{ijk}$.

C. CAUSALITY NETWORK PREDICTION

Causality Network prediction or inference is to calculate the probability of the hypothesis variables at certain status according to the network topology and conditional probability distribution of the evidence variable. An inference or query $P(Q = q | E = e_0)$ is to calculate the posterior probability of a query variable Q being at its specific value q in the condition of given evidence e_0 for node E.

There are many existing network inference algorithms, such as variable elimination algorithm, and junction tree algorithm (JT). These algorithms utilize the hypothesis variables and specific independence relations induced by evidence in the Bayesian network to simplify the updating task. JT implements the inference procedure in four steps [37],

1. Cluster the nodes into several cliques,
2. Connect the cliques to form a junction tree,
3. Propagate information in the network,
4. Answer a query.

The inference starts from a root clique. The core step of message propagation consists of a message collection phase and a distribution phase. The cliques of the junction tree are connected by separators such that the so-called junction tree property holds. When a message is passed from one clique X to another clique Y, it is mediated by the sepset S between the two cliques. Every conditional probability distribution of the original Bayesian network is associated with a clique such that the domain of the distribution is a subset of the clique domain(we use the notation $dom(\phi)$ to refer to the domain of a potential ϕ). The set of distributions Φ_X associated with a clique X are in standard junction tree architectures combined to form the initial clique X.

$$\phi_X = \prod_{\phi \epsilon \Phi_X} \phi. \quad (16)$$

For a clique, a potential or a message is a mapping from the value assignments of the nodes to the set [0, 1.0]. A message pass from X to Y occurs with two procedures: projection and absorption based on the HUGIN architecture [38]. The projection procedure saves the current potential and assigns a new one to S:

$$\phi_S^{old} \leftarrow \phi_S, \text{ and } \phi_S \leftarrow \sum_{X \setminus S} \phi_X. \tag{17}$$

The absorption procedure assigns a new potential to Y using both the old and the new tables of S:

$$\phi_Y \leftarrow \phi_Y \frac{\phi_S}{\phi_S^{old}}. \tag{18}$$

where ϕ_S is the current separator potential, ϕ_S^{old} is the old separator potential, ϕ_X is the clique potential for X, ϕ_Y is the clique potential for Y.

The query answering step has two procedures. First, the marginalization procedure calculates the joint probability of Q and $E = e_0$: $P(Q, E = e_0) = \sum_{X\{Q\}} \phi_X$. Second, the normalization procedure calculates the inference result:

$$P(Q = q | E = e_0) = \frac{P(Q = q, E = e_0)}{\sum_Q P(Q, E = e_0)}. \tag{19}$$

The fault of operational variables is an intervention that have various effect on the production process. The main task in causal inference is to predict the system output under the faulty intervention.

III. SIMULATION RESULTS

In order to evaluate the performance of the proposed method, we report the experiment results from three aspects: the causal direction identification of multi-variables, network parameter learning and probability inference.

A. PUBLIC DATA SETS EXAMPLE

Four published data sets proposed by Mooij and Janzing [39] is used to test the effectiveness of the nonlinear multivariate causal model. The cause-effect pairs are available at http://webdav.tuebingen.mpg.de/cause-effect/. This database with different data is considered as the benchmark for testing causal detection algorithms. Data set (1) contains the ground altitude and temperature sampled at 349 stations, US. Data set (2) is census income dataset which contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau. The variables include age and wage per hour. Data set (3) gives the attribute information (age and heart rate) from Cardiac Arrhythmia database. Data set (4) include the population with sustainable access to improved drinking water sources (%) total, and the infant mortality rate (per 1000 live births) both sexes, 2006. These four data sets have different attributes, which is sufficient to show the general and comprehensive nature of the data. Figure 3 gives the scatterplots of the selected data set (1-4). Table 1 summaries the causal results obtained by the multivariate causality model.

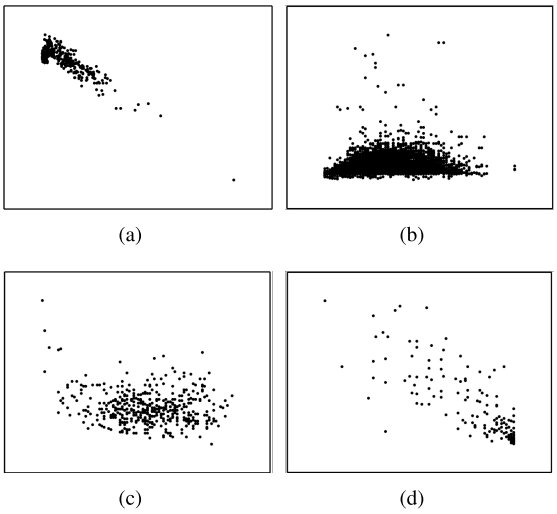


FIGURE 3. Scatterplots of the data sets, (a), (b), (c), (d) correspond to data 1, 2,3 4, respectively.

TABLE 1. The causal result of the public datasets.

Data Sets	#1	#2	#3	#4
Information for data	x:altitude y:temperature	x:age y:wage per hour	x:age y:heart rate	x:population y:infant mortality rate
Real direction	$x \to y$	$x \to y$	$x \to y$	$x \to y$
Test results	$x \to y$	$x \to y$	$x \to y$	$x \to y$
True or false	true	true	true	true

TABLE 2. Results of independence test under different assumed causal directions.

causal directions	$x \to y$	$y \to x$
#1	1.7×10^{-3}	6.5×10^{-3}
#2	1.2×10^{-4}	6.7×10^{-4}
#3	3.5×10^{-3}	8.1×10^{-3}
#4	2.2×10^{-3}	5.7×10^{-3}

Table 2 shows the results of independence test on x and y for Data sets (1-4) under different assumed causal directions. The statistics under different causal direction assumptions are calculated separately.

Comparing the test statistics under two assumed causal directions in Table 2, the causal direction of each set all are determined as $x \to y$, which is consistent to the real causal relationship. We can conclude that the proposed method can correctly identify the causal direction regardless the diversity of data.

B. TE PROCESS

1) TE PROCESS CAUSAL STRUCTURE

In order to illustrate the applicability of the proposed method in the actual complex industrial process, we establish the network topology of TE process and predict the alarm variables. TE platform simulates an actual chemical process, as shown in Figure 4, which is widely used as a benchmark to test the process control and fault diagnosis technology. TE process

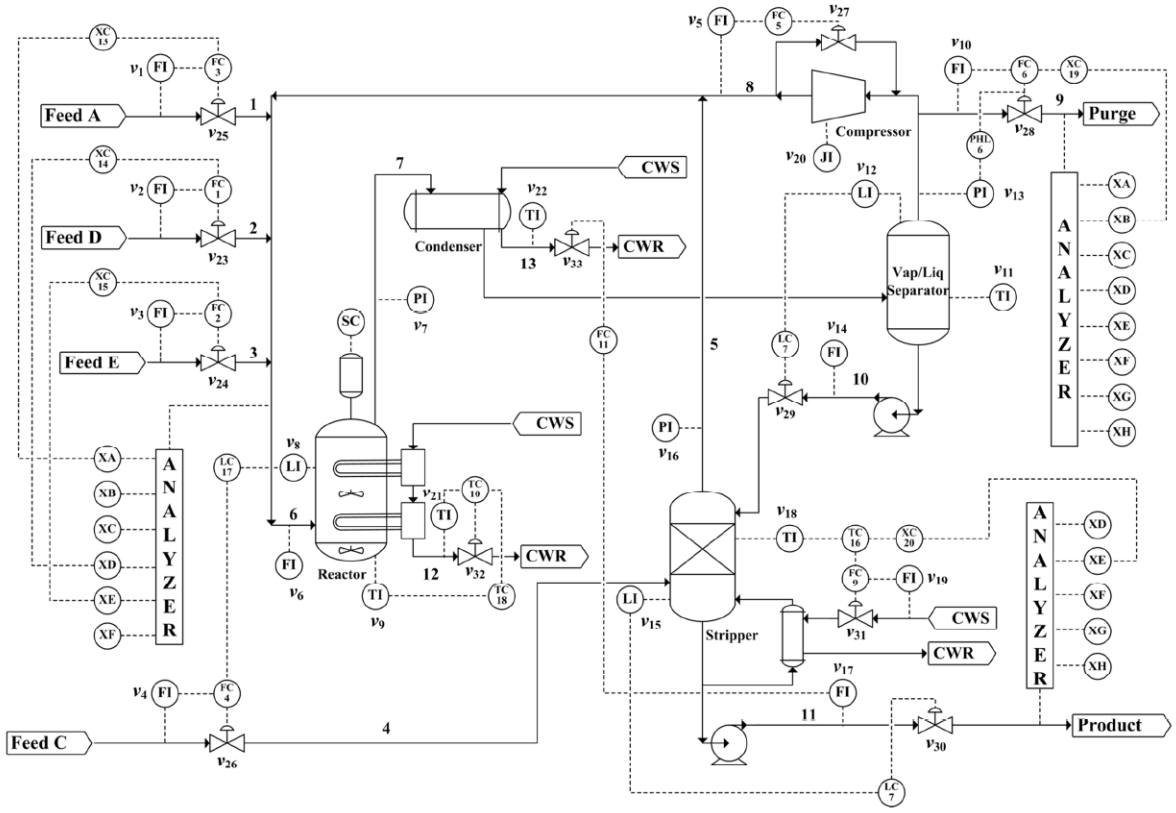


FIGURE 4. TE industrial process flow diagram.

TABLE 3. Forecasted variables in the TE process.

variable	Physical meaning	units
X_1	Recycle flow	km^3/h
X_2	Reactor feed rate	km^3/h
X_3	Reactor pressure	kPa
X_4	Reactor level	%
X_5	Reactor temperature	°C
X_6	Product separator level	%
X_7	Compress work	KW
X_8	Reactor cooling water outlet temperature	°C

consists of five major unit operations, including a reactor, a product condenser, a vapour-liquid separator, a recycle compressor, and a product gas stripper. First, the gaseous feed is converted to a liquid product by reaction in the reactor, where the condensate is contained in the reactor to absorb the heat generated by the reaction, and the product with some unreacted material mixed is discharged in gaseous form.

TE process has 12 manipulated variables, 22 continuous measurements, and 19 composition measurements. TE system generates various alarm information under several predefined specific faults. Table 3 lists 8 process variables to forecast their alarm status.

From the mechanism analysis of TE process, we know that when the feed X_2 increases, the material is first entered into the reactor, so the reactor level X_4 must increase. So the reactor feed X_2 directly affects the reactor level X_4. The temperature of cooling water X_8 and the feed of the reactor X_2 are the main causes of the reactor temperature X_5. The reactor pressure X_3 is synchronized with the reactor temperature X_5 changes according to the general physical principle. In addition, once the reaction in the reactor is more intense, the compressor module power X_7 will be synchronized to strengthen due to the sequential loop. At the same time, the reactor pressure X_3 will act on the recovered flow X_1 and the material level X_6 in the separator. We can determine the initial structure of the causality network, Bnet0 shown in Figure 5, by the expert prior knowledge and the sample data correlation analysis.

The pre-defined fault is random variations in A, B, C compositions in stream 4. The corresponding data of eight variables are collected from the simulation platform. The reaction length is 700 hours in order to ensure that the data is sufficient to reflect the system process. We obtain 500 sets of sampling data after the equal time decimate. The causal direction of the paired variables is shown in Table 4. The directed acyclic graph, Bnet1, is constructed using the proposed causal method and the corresponding data, which is shown in Figure 6(a). An alternative network, Bnet2, are shown in Figure 6(b), which is obtained from the traditional BN structure learning method-K2 algorithm [40] which needs

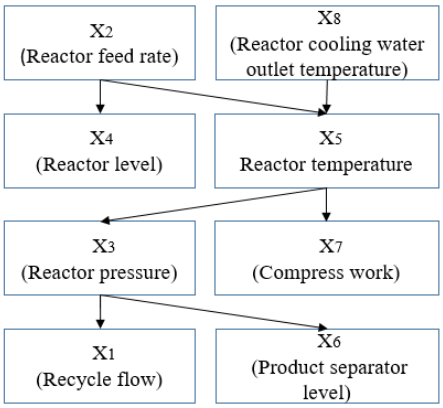


FIGURE 5. The network Bnet0 from the mechanism analysis.

TABLE 4. Causal direction of TE variables.

Information of variables	Statistic(positive/ reverse)	Causal direction
X_2:Reactor feed rate X_5:Reactor temperature	$5.7 \times 10^{-6}/8.2 \times 10^{-6}$	$X_2 \rightarrow X_5$
X_5:Reactor temperature X_8:Reactor cooling water outlet temperature	$7.1 \times 10^{-6}/2.9 \times 10^{-6}$	$X_8 \rightarrow X_5$
X_2:Reactor feed rate X_4:Reactor level	$3.4 \times 10^{-4}/8.5 \times 10^{-4}$	$X2 \rightarrow X_4$
X_5:Reactor temperature X_7:Compress work	$7.3 \times 10^{-4}/9.2 \times 10^{-4}$	$X_5 \rightarrow X_7$
X_3:Reactor pressure X_5:Reactor temperature	$7.6 \times 10^{-5}/4.5 \times 10^{-5}$	$X_5 \rightarrow X_3$
X_3:Reactor pressure X_6:Product separator level	$2.9 \times 10^{-6}/3.9 \times 10^{-6}$	$X_3 \rightarrow X_6$
X_1:Recycle flow X_3:Reactor pressure	$6.6 \times 10^{-6}/2.7 \times 10^{-6}$	$X_3 \rightarrow X_1$

to set the node order. Figure 6(c) shows the network structure Bnet3 learned with the expectation maximization (EM) algorithm.

Comparing the process analysis structure Bnet0 and Bnet1 determined by the proposed causal method, it is seen that Bnet1 is exactly consistent to Bnet0. The structure determined using the proposed method exactly matches the mechanism and expert knowledge, which indicates that the causal structure is credible and accurate. However, Bnet2 and Bnet3 learned from the traditional BN methods is not consistent with the mechanism and shows a big gap from the actual conditions. This demonstrates that the general BN learning method fails when applied to the complex nonlinear systems, while the proposed multivariate causality model proves its superiority.

2) TE PARAMETER LEARNING

Once the TE network structure is determined, the alarm prediction model can be obtained by learning the network parameters. In general, the process alarm event can be divided into five alarm levels, namely, high_high alarm (HH), high alarm(H), normal(N), low alarm(L) and low_low alarm(LL), corresponding to the number 1,2,3,4,5. The first step is to

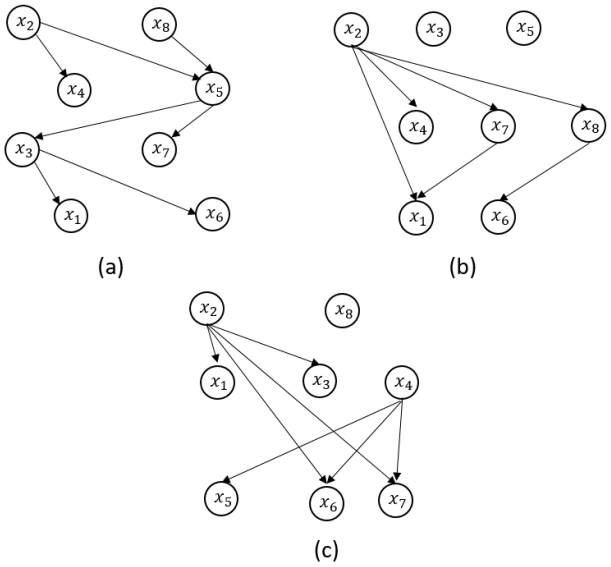


FIGURE 6. The network compare: (a) Bnet1, (b) Bnet2, (c) Bnet3.

discretize the continuous variables into five alarm levels by setting different thresholds, shown in Table 5.

Here we adopt the MLE algorithm to learn the network parameters and get a complete probability table. Suppose that the initial probability of the alarm level in the normal condition is theoretically divided equally. Then we can get the conditional probability tables for all variables based on Bayesian network parameter learning. Considering two root nodes X_2 and X_8, their corresponding probabilities for five status are 0.0843, 0.2211, 0.4704, 0.2026 and 0.0217, respectively. The probability of other descendant variables as shown in Figure 7. We use hot plot to show the probability since the precise value has nothing meaning for alarm prediction and inference. The color represents the probability range between 0 and 1.

We are concerned about the probability of close to 1, because this is the key point in determining the inference results. When the probability is less than 0.5, the result situation will not likely appear in the actual inference. Figure 7(a) shows the probability of X_5 under the combined action of X_2 and X_8. The abscissa is the state condition of X_8 and X_2, and the ordinate is the color corresponding to the probability value of X5. $P(X_5 = 1|X_8 = 1, 2 \text{ and } X_2 = 1) \approx 1$ in the lower left corner of (a). It means that X5 occurs the low_low alarm with the probability close to 1 when X_2 and X_8 are in the low_low alarm state. $P(X_5 = 5|X_8 = 4, 5 \text{ and } X2 = 5) \approx 1$ in the upper right corner of (a) means that X5 occurs the high_high alarm with the probability close to 1 when X_2 and X_8 are in the high_high alarm state. These inference results are consistent with the actual mechanism.

Figure (b)-(e) reflect the probability relationship between bivariate variables. Figure 7(b) shows the probability of X_4 under the action of X_3. $P(X_4 = 5|X_3 = 5) \approx 1$ in the upper right corner of (b) means that the probability of X_4 occurs the

TABLE 5. The range of thresholds for alarm variables in different states.

Alarm status	$X_1(km^3/h)$	$X_2(km^3/h)$	X_3(kPa)	$X_4(\%)$	$X_5(^\circ C)$	$X_6(\%)$	X_7(KW)	$X_8(^\circ C)$
1	<31	<46	<2789	<62.5	<122.7	<45	<268	<102.25
2	31-32	46-47	2789-2796	62.5-63.8	122.7-122.87	45-47.2	268-272.3	102.25-102.41
3	32-33	47-48.3	2796-2802	63.8-66	122.87-122.93	47.2-52.2	272.3-274	102.41-102.55
4	33-34	48.3-49.5	2804-2809	66-66.8	122.93-123.2	52.2-53	274-280	102.55-102.7
5	>34	>49.5	>2809	>66.8	>123.2	>53	>280	>102.7

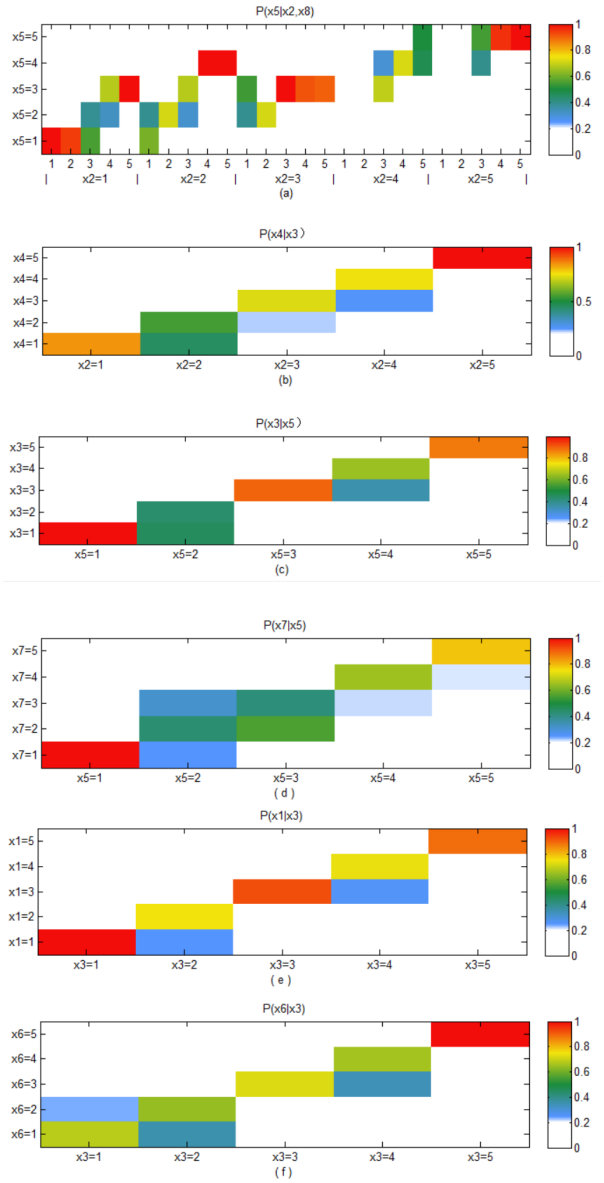


FIGURE 7. Conditional probability of the descendant variables: (a) $P(X_5|X_8, X_2)$, (b)$P(X_4|X_3)$, (c) $P(X_3|X_5)$, (d) $P(X_7|X_5)$, (e) $P(X_1|X_3)$, (f) $P(X_6|X_3)$.

high_high alarm close to 1 when X_3 in the high_high alarm state. However, $P(X_4 = 1|X_3 = 5) = 0$ in the lower right corner of (b) means that X_4 occurs the high_high alarm with the probability close to 0 when X3 in the low_low alarm state. $P(X_4 = 1 \text{and} X_4 = 2|X_3 = 2) \approx 0.5$ in the green area of (b) means the probability of X4 occurs the low alarm or low_low alarm almost same when X3 in the low alarm state. Similarly, we can analyze the results consistent with the mechanism in Figure 7(c)-(e).

TABLE 6. Alarm level prediction of compress work X7.

Number	X_2	X_8	X_5	X_7 real status	X_7 Prediction	Maximum probability
1	1	2	1	2	1	0.4571
2	2	1	2	1	1	0.6501
3	1	2	2	2	2	0.7627
4	2	1	2	2	2	0.6729
5	1	2	2	1	1	0.6896
6	3	3	2	3	1	0.8760
7	3	3	2	3	3	0.6344
8	3	3	3	2	2	0.8563
9	3	3	2	3	2	0.3454
10	2	3	3	3	3	0.5073
11	3	3	3	2	3	0.4432
12	3	2	3	3	3	0.5696
13	4	3	4	4	3	0.3128
14	3	4	4	4	4	0.6284
15	4	5	5	5	5	0.7557
16	4	3	4	4	5	0.3783
17	5	5	4	4	4	0.7947
18	4	5	4	4	4	0.8325
19	5	4	5	4	5	0.6454
20	5	4	4	5	5	0.8113

3) TE ALARM PREDICTION

Alarm prediction is a top-down inference according to the evidences inference conclusion. The probabilistic analysis calculates the likelihood of each status for the result variable may occur. The discrete status corresponding to the maximum probability is the alarm prediction result.

Using the established multivariate causality network model, compress work X_7 is predicted when its parent variables X_2, X_8 and X_5 are known. The prediction results for model Bnet1 are shown in Table 6.

The total prediction accuracy for the 20 simulation experiments is 75%. When the maximum probability of the predicted value is greater than 0.5, the prediction result is confident. Furthermore, the predictions with a high probability is consistent with the true status. When the maximum probability of the predicted value is less than 0.5, the prediction result is not believable and accurate. The mispredictions confuse the adjacent status, such as the normal status 2 and Low alarm 3 (or high alarm 2). The simulation results show that the multivariate causality network can find the intrinsic relationships among various process variables, and give precise fault or alarm prediction.

IV. CONCLUSION

We propose a multivariate causality model to analyze the causal direction of multi-variable and final determine the network topology. The proposed method can describe the system structure more accurate than the traditional BN structure learning method, when the industrial process is high complex. Combined with network parameters learning and evidence inference technique, we can accurately monitor the industrial process. The validity of the proposed method is verified via the public data and TE process. We obtain a compact variable network and confident alarm prediction of the TE process based on causal analysis and probability inference. Both the methodology and the simulation results show that our research results have great value for the process industry modeling and monitoring.

There are some issues worth further discussion. The computing efficiency of the proposed multivariate post-nonlinear acyclic causal modeling method should be considered when solving the large-scale real world causal analysis problems. Developing efficient algorithm for causal discovery of multiple variables based on the general functional causal models is still an important topic. To make the causal discovery computationally efficient, one may have to limit the complexity of the causal structure, such as decrease the number of direct causes of each variable. So far, a smart optimization procedure instead of exhaustive search is still missing in the literatures.

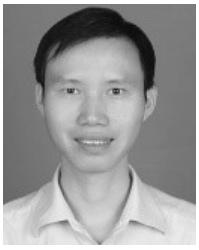
REFERENCES

[1] W. Wolf, "Cyber-physical systems," *Computer*, vol. 42, no. 3, pp. 88–89, Mar. 2009.
[2] G. Sheng *et al.*, "Mathematical models for simulating coded digital communication: a comprehensive tutorial by big data analytics in cyber-physical systems," *IEEE Access*, vol. 4, pp. 9018–9026, Dec. 2016.
[3] I. Friedberg, X. Hong, K. Mclaughlin, P. Smith, and P. C. Miller, "Evidential network modeling for cyber-physical system state inference," *IEEE Access*, vol. 5, pp. 17149–17164, 2017.
[4] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, 2013.
[5] Y. Wang, D. Zhao, Y. Li, and S. X. Ding, "Unbiased minimum variance fault and state estimation for linear discrete time-varying two-dimensional systems," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 5463–5469, Oct. 2017.
[6] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometrics Intell. Lab. Syst.*, vol. 171, pp. 16–25, Dec. 2017, doi: 10.1016/j.chemolab.2017.09.021.
[7] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.
[8] W. Ge, J. Wang, J. Zhou, H. Wu, and Q. Jin, "Incipient fault detection based on fault extraction and residual evaluation," *Ind. Eng. Chem. Res.*, vol. 54, no. 14, pp. 3664–3677, Apr. 2015.
[9] Z. Lou and Y. Wang, "Multimode continuous processes monitoring based on hidden semi-Markov model and principal component analysis," *Ind. Eng. Chem. Res.*, vol. 56, no. 46, pp. 13800–13811, 2017.
[10] R. Wang, J. Wang, J. Zhou, and H. Wu, "Fault diagnosis based on the integration of exponential discriminant analysis and local linear embedding," *Can. J. Chem. Eng.*, vol. 96, no. 2, pp. 463–483, 2018, doi: 10.1002/cjce.22921.
[11] J. Wang, J. Zhang, B. Qu, H. Wu, and J. Zhou, "Unified architecture of active fault detection and partial active fault-tolerant control for incipient faults," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1688–1700, Jul. 2017.
[12] Z. Lou, D. Shen, and Y. Wang, "Two-step principal component analysis for dynamic processes monitoring," *Can. J. Chem. Eng.*, vol. 96, no. 1, pp. 160–170, 2018, doi: 10.1002/cjce.22855.
[13] Q. Jiang, X. Yan, and J. Li, "PCA-ICA integrated with Bayesian method for non-Gaussian fault diagnosis," *Ind. Eng. Chem. Res.*, vol. 55, no. 17, pp. 4979–4986, May 2016.
[14] Z. Ge, L. Xie, U. Kruger, and Z. Song, "Local ICA for multivariate statistical fault diagnosis in systems with unknown signal and error distributions," *AIChE J.*, vol. 58, pp. 2357–2372, Aug. 2012.
[15] J. Yu, "Localized Fisher discriminant analysis based complex chemical process monitoring," *AIChE J.*, vol. 57, no. 7, pp. 1817–1828, Jul. 2011.
[16] S. He, Y. Wang, and C. Liu, "Modified partial least square for diagnosing key-performance-indicator-related faults," *Can. J. Chem. Eng.*, vol. 96, no. 2, pp. 444–454, 2018, doi: 10.1002/cjce.23002.
[17] K. W. Hipel, K. D. Marc, and L. Fang, *The Graph Model for Conflict Resolution*. Hoboken, NJ, USA: Wiley, 2011, pp. 507–520. [Online]. Available: http://www.eolss.net/Sample-Chapters/C14/E1-40-04-01.pdf
[18] X. Li *et al.*, "DeepSaliency: Multi-task deep neural network model for salient object detection," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3919–3930, Aug. 2016, doi: 10.1109/TIP.2016.2579306.
[19] Y. Jiang, Z. Deng, F.-L. Chung, and S. Wang, "Multi-task TSK fuzzy system modeling using inter-task correlation information," *Inf. Sci.*, vol. 298, pp. 512–533, Mar. 2015, doi: 10.1016/j.ins.2014.12.007.
[20] A. Das, "Theoretical and experimental comparison of off-grid sparse Bayesian direction-of-arrival estimation algorithms," *IEEE Access*, vol. 5, pp. 18075–18087, Aug. 2017.
[21] C. Xu, C. Gao, Z. Zhou, Z. Chang, and Y. Jia, "Social network-based content delivery in device-to-device underlay cellular networks using matching theory," *IEEE Access*, vol. 5, pp. 924–937, 2017.
[22] J. Zhu, Z. Ge, Z. Song, L. Zhou, and G. Chen, "Large-scale plant-wide process modeling and hierarchical monitoring: A distributed Bayesian network approach," *J. Process Control*, vol.13, no. 4, pp. 1877–1885, Aug. 2017.
[23] X. Sun, D. Janzing, and B. Scholkopf. "Causal inference by choosing graphs with most plausible Markov kernels," presented at the 9th Int. Symp. Artif. Intell. Math., Fort Lauderdale, FL, USA, 2006.
[24] A. Hyvärinen and S. M. Smith, "Pairwise likelihood ratios for estimation of non-Gaussian structural equation models," *J. Mach. Learn. Res.*, vol. 14, pp. 111–152, Jan. 2010.
[25] G. Mai, Y. Hong, S. Peng, and Y. Peng, "Inferring causal direction from multi-dimensional causal networks for assessing harmful factors in security analysis," *IEEE Access*, vol. 5, pp. 20009–20019, 2017, doi: 10.1109/ACCESS.2017.2746539.
[26] K. Zhang and A. Hyvärinen, "Distinguishing causes from effects using nonlinear acyclic causal models," in *Proc. JMLR Workshop Conf.*, vol. 6. 2010, pp. 157–164.
[27] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen "A linear non-Gaussian acyclic model for causal discovery," *J. Mach. Learn. Res.*, vol. 7, no. 4, pp. 2003–2030, 2006.
[28] S. Shimizu *et al.*, "DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model," *J. Mach. Learn. Res.*, vol. 12, no. 2, pp. 1225–1248, Apr. 2011.
[29] J. Peters, J. Mooij, and D. Janzing, "Identifiability of causal graphs using functional models," presented at the Conf. Uncertainty Artif. Intell., Arlington, TX, USA, 2011.
[30] W.-K. Chen, *Statistics and Causality: Methods for Applied Empirical Research*. Hoboken, NJ, USA: Wiley, 2016. sec. 8, pp. 123–135. [Online]. Available: http://download.e-bookshelf.de/download
[31] Z. Hao and L. Hongguang, "Optimization of process alarm thresholds: A multidimensional kernel density estimation approach," *Process Safety Progr.*, vol. 33, no. 3, pp. 292–298, 2015.
[32] J. M. Górriz *et al.*, "Case-based statistical learning: A non-parametric implementation with a conditional-error rate SVM," *IEEE Access*, vol. 5, no. 2, pp. 11468–11478, 2017.
[33] K. B. Korb and A. E. Nicholson. *Bayesian Artificial Intelligence*. Boca Raton, FL, USA: CRC Press, 2004, pp. 15–64.
[34] A. Mehmood, A. Khanan, A. H. H. M. Mohamed, and H. Song, "ANTSC: An intelligent Naïve Bayesian probabilistic estimation practice for traffic flow to form stable clustering in VANET," *IEEE Access*, to be published, doi: 10.1109/ACCESS.2017.2732727.
[35] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000.

[36] A. Gretton, K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. J. Smola, "A kernel statistical test of independence," presented at the Adv. Neural Inf. Process. Syst. Conf., Cambridge, MA, USA, 2008.
[37] M. Borsotto, W. Zhang, E. Kapanci, A. Pfeffer, and C. Crick, "A junction tree propagation algorithm for Bayesian networks with second-order uncertainties," presented at the Int. Conf. Tools Artif. Intell. (ICTAI), Arlington, VA, USA, 2006.
[38] F. V. Jensen, S. L. Lauritzen, and K. G. Olesen, "Bayesian updating in causal probabilistic networks by local computations," *Comput. Statist. Quaterly*, vol. 4, pp. 269–282, Jun. 1990.
[39] J. Mooij and D. Janzing. (Oct. 2008). *Distinguishing Between Cause and Effect*. [Online]. Available: http://www.kyb.tuebingen.mpg.de/bs/people/jorism/causality-data/
[40] B. Lerner and R. Malka, "Investigation of the K2 algorithm in learning Bayesian network classifiers," *Appl. Artif. Intell.*, vol. 25, no. 1, pp. 74–96, Jan. 2011.

XIAOLU CHEN is currently pursuing the Ph.D. degree in control science and engineering with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. Her research interests include the modeling and fault diagnosis of complex industrial processes, data causality analysis, and intelligent learning algorithm.

JING WANG (M'17) received the B.S. degree in industry automation, the Ph.D. degree in control theory and control engineering from the Northeastern University, in 1994 and 1998, respectively. She was a Visiting Professor at University of Delaware, USA, in 2014. She is currently a Professor with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. Her research interests are oriented to different aspects, including application of advanced control schemes to nonlinear, multivariable, constrained industrial processes, modeling, optimization and control for complex industrial process, nonlinear model based control of polymer microscopic quality in chemical reactor, process monitoring, and fault diagnosis for complex industrial process.



JINGLIN ZHOU (M'14) received the B.Eng., M.Sc., and Ph.D. degrees from the Daqing Petroleum Institute, Hunan University, Changsha, China, and the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1999, 2002, and 2005, respectively. He was an Academic Visitor with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, U.K. He is currently a Professor with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing. His current research interests include stochastic distribution control, fault detection and diagnosis, and variable structure control and their applications.