Chapter 13
Bayesian Causal Network for Discrete Variables

Ensuring the safety of industrial systems requires not only detecting the faults, but also locating them so that they can be eliminated. The previous chapters have discussed the fault detection and identification methods. Fault traceability is also an important issue in industrial system. This chapter and Chap. 14 aim at the fault inference and root tracking based on the probabilistic graphical model. This model explores the internal linkages of system variables quantitatively and qualitatively, so it avoids the bottleneck of multivariate statistical model without clear mechanism. The exacted features or principle components of multivariate statistical model are linear or nonlinear combinations of system variables and have not any physical meaning. So the multivariate statistical model is good at fault detection and identification, but not at fault root tracking.

Bayesian network (BN) can estimate and predict the potentially harmful factors of the general system, but its structure learning has some deficiencies when it is applied to the complex 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. Usually, a generative model (linear or nonlinear) is built to explain the data generating process, i.e., the causalities. A variety of causal discovery methods have been proposed recently to find the causalities (Hyvärinen et al. 2010; Hong et al. 2017). The most classical method is the linear non-gaussian acyclic model (LiNGAM) (Shimizu et al. 2010), in which the full structure of BN is identifiable without pre-specifying a causal order of the variables. The improved LiNGAM method is proposed to estimate the causal order of variables without any prior structure knowledge and provide better statistical performance (Shimizu et al. 2011). The nonlinear causality of a pair of variables is discovered in Johnson and Bhattacharyya (2015), where the proposed method shows a limitation when dealing with the multivariate variables.

The above 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 have been proposed over the last few years, their practical performance has not been studied systematically. These methods have yet

to be applied to the actual industrial systems which usually do not meet the linear and bivariate assumptions. To address the above issues, this chapter proposes a more generalized multivariate post-nonlinear acyclic causal model for the complex industrial process. The proposed multivariate post-nonlinear acyclic causal model, named as Bayesian Causal Network (BCN), can easily find the multi-variables causality. It shows more compact structure and consistency with mechanism, compared with the traditional BN structure. In addition, it avoids the complex learning mechanism of traditional BN, so is easier to implement without compromising accuracy.

13.1 Construction of Bayesian Causal Network

It is known that there are many ways to describe the system characteristic according to the observational data and expert knowledge, such as graph model (Hipel et al. 2011), neural network model (Li et al. 2016), fuzzy model (Jiang et al. 2015). 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, a typical graph model, is the main method to deal with the knowledge representations and uncertainties based on the probability theory. It builds the causality and probability within the process components and the system variables from the prior knowledge and process data. BN consists of the structure learning and the parameter learning, in which the structure learning aims at determining the causalities within system variables and the parameter learning aims at revealing the quantitative relationship of these causalities. Bayesian network has been applied to fault diagnosis, financial analysis, automatic target recognition, military, and many other areas (Zhu et al. 2017).

13.1.1 Description of Bayesian Network

Bayesian network, also known as Belief Network or directed acyclic graphical model, is a probabilistic graphical model. It first proposed by Judea Pearl in 1985 (Pearl 1986). It is an uncertainty processing model that simulates the causal relationship in the human reasoning process, and its network topology is a directed acyclic graph (DAG). The nodes in the directed acyclic graph represent the random variables, including the observable variables, hidden variables, unknown parameters, etc. Variables or propositions that are believed to have a causal relationship (or non-conditional independence) are connected by arrows (in other words, the arrow connecting two nodes represents whether the two random variables have a causal relationship or are not conditionally independent). If two nodes are connected by a single arrow, it means that one of the nodes is "cause" and the other is "effect", a conditional probability value is used to describe the causality degree quantitatively.

Fig. 13.1 Bayesian network
example
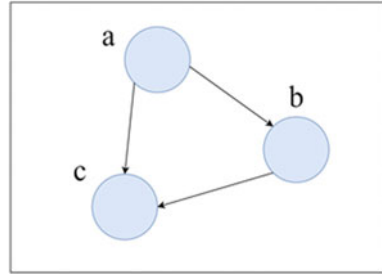
For example, assume that node A directly affects node B, then $A \rightarrow B$. The arrow from A to B is used to establish a directed arc (A, B) from node A to node B, and the weight (its connection strength) is determined by the conditional probability $P(B|A)$. In short, a BN is formed by drawing the random variables in a directed graph according to whether they are conditionally independent. It usually uses circle to represent the random variables (nodes) and arrow to represent the conditional dependencies. Figure 13.1 gives a simple Bayesian network (Ishak et al. 2011).

13.1.2 Establishing Multivariate 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 identified such that it could distinguish the causes from the effects. A nonlinear and multivariable system always possesses the following three characteristics (Chen et al. 2018):

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 complex industrial systems, 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 DAG to represent the relationship among multiple observed variables. Mathematically, the generating process of X_i is

$$X_i = f_{i,2}(f_{i,1}(\boldsymbol{PA}_i) + \boldsymbol{e}_i), \tag{13.1}$$

where the observed variables $X_i, i = \{1, 2, \dots, n\}$ are arranged in a causal order, such that no later variable causes any earlier variable. $\boldsymbol{PA}_i$ is the direct cause of

X_i. $f_{i,1}$ denotes the nonlinear effect of this cause, and $f_{i,2}$ denotes the 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 (13.1) satisfies the aforementioned three characteristics: 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 $f_{i,2}$ reflects the nonlinear distortion caused by the sensor or measurement.

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

$$X_j = f_{j,2}(f_{j,1}(X_i) + e_j), \tag{13.2}$$

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 $X_i \rightarrow X_j$ as follows:

$$\begin{aligned} X_i &= f_{j,1}^{-1}(s_i), \\ X_j &= f_{j,2}(s_i + s_j). \end{aligned} \tag{13.3}$$

X_i and X_j in (13.3) are post-nonlinear (PNL) mixtures of independent sources s_i and s_j. The PNL mixing model can be seen as a special case of the general nonlinear independent component analysis (ICA) model. Here we use nonlinear ICA method to solve this problem (13.3).

Generally there are two possibility to describe the causal relation between any two random variables X_i and X_j, $(X_i \rightarrow X_j$ or $X_j \rightarrow X_i)$. We should identify the correct relation by judging which one satisfies the assumed model (13.2). If the causal relation is $X_i \rightarrow X_j$ (i.e., X_i and X_j satisfy the model (13.2)), we can invert the data generating process (13.2) to recover the disturbance e_j, which is expected to be independent from X_i. Two steps are used to examine the possible causal relationships between variables.

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

$$e_j = f_{j,2}^{-1}(X_j) - f_{j,1}(X_i), \tag{13.4}$$

where e_j is independent from X_i. Thus perform nonlinear ICA using the structure in Fig. 13.2 and the outputs of system are

$$\begin{aligned} Y_i &= X_i, \\ Y_j &= e_j = g_j(X_j) - g_i(X_i)). \end{aligned} \tag{13.5}$$

Fig. 13.2 Constrained
nonlinear ICA system used
to verify if the causal relation
$X_i \to X_j$ holds
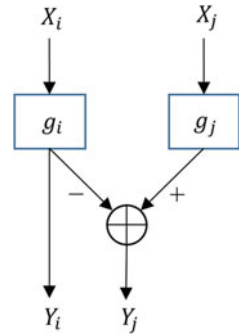
The nonlinearities g_i and g_j is modeled by Multi-layer perceptrons (MLP's), and the parameters in g_i and g_j are learned by making Y_i and Y_j as independent as possible, i.e., minimizing the mutual information between Y_i and Y_j,

$$I(Y_i, Y_j) = H(Y_i) + H(Y_j) - H(Y), \qquad (13.6)$$

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

$$
\begin{aligned}
H(Y) &= -\mathbb{E}\left[\log p_Y(Y)\right] \\
&= -\mathbb{E}\left[\log p_Y(X) - \log |J|\right] \\
&= H(X) + \mathbb{E}\left[\log |J|\right].
\end{aligned}
\qquad (13.7)
$$

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.,

$$
\begin{aligned}
J &= \frac{\partial(Y_i, Y_j)}{\partial(X_i, X_j)}, \\
|J| &= \begin{vmatrix} 1 & 0 \\ g_i' & g_j' \end{vmatrix} = |g_j'|.
\end{aligned}
\qquad (13.8)
$$

Substitute (13.7) and (13.8) into (13.6), we have

$$I(Y_i, Y_j) = H(Y_i) + H(Y_j) - \mathbb{E}[\log |J|] - H(X) \qquad (13.9)$$

$$= -\mathbb{E}\left[\log p_{Y_i}(Y_i)\right] - \mathbb{E}\left[\log p_{Y_j}(Y_j)\right] - \mathbb{E}\left[\log |g_j'|\right] - H(X), \qquad (13.10)$$

where $H(X)$ does not depend on the parameters in g_i and g_j and can be considered as constant. The minimization problem (13.10) 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 test. The kernel-based

statistical test is adopted with the significance level $= 0.01$ (Giga 2014). Denote the test statistic as $test_{i \rightarrow j}$. If $test_{i \rightarrow j} > test_{j \rightarrow i}$, it indicates that Y_i and Y_j are not independent, that is $X_i \rightarrow X_j$ does not hold. Repeat the above procedure with X_i and X_j exchanged to verify if $X_j \rightarrow X_i$ holds. If $test_{i \rightarrow j} < test_{j \rightarrow i}$, it concludes 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 system, there are n process variables. Following a test sequence, $X_1 \rightarrow X_2, X_1 \rightarrow X_3, \ldots, X_{n-1} \rightarrow X_n$, the N group statistics should be tested,

$$N = n + (n-1) + (n-2) + \cdots + 1 = \frac{n(n-1)}{2}. \qquad (13.11)$$

The total computation is in direct proportion to $2 \times N$. 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

$$\begin{aligned} \boldsymbol{A} &= [test_{X_1 \rightarrow X_2}, test_{X_1 \rightarrow X_3}, \ldots, test_{X_{n-1} \rightarrow X_n}], \\ \boldsymbol{B} &= [test_{X_2 \rightarrow X_1}, test_{X_3 \rightarrow X_1}, \ldots, test_{X_n \rightarrow X_{n-1}}]. \end{aligned} \qquad (13.12)$$

Comparing the corresponding elements of the vectors $\boldsymbol{A}$ and $\boldsymbol{B}$, the causal direction of this pair of variables is determined according to the smaller statistic. Once the causality of all variables is found based on the above cyclic search, integrate them into a DAG.

13.1.3 Network Parameter Learning

The multivariate causality model gives a framework similar to the Bayesian network to find the internal structure of the complex systems. Its 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 $\boldsymbol{U} = \{X_1, \ldots, X_n\}$ 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 $\boldsymbol{U}$. Formally, the Bayesian network for $\boldsymbol{U}$ is constructed as a pair $\boldsymbol{B} = < \boldsymbol{G}, \boldsymbol{\Theta} >$. $\boldsymbol{G}$ is a directed acyclic graph whose vertices is correspond to the random variables $X_1, \ldots, X_n$. $\boldsymbol{\Theta}$ 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 $\boldsymbol{pa}_i^j$ is one of components in the complete parent set $\boldsymbol{PA}_i$ of X_i in $\boldsymbol{G}$. Every variable X_i is conditionally independent of its non-descendants given its parents (Markov condition). The joint probability distribution over set $\boldsymbol{U}$ is

$$P_B(X_1, \ldots, X_n) = \prod_{i=1}^{n} P_B(X_i | \boldsymbol{PA}_i) = \prod_{i=1}^{n} \theta_{X_i | \prod \boldsymbol{PA}_i}. \tag{13.13}$$

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

Give a data set $D = \{D_1, \ldots, D_N\}$ of all Bayesian network nodes. The goal of parameter learning is to find the most probable values for $\boldsymbol{\Theta}$. These values best explain the data set 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, we have

$$L_D(\theta) = \log \prod_{i=1}^{n} \prod_{j=1}^{q_i} \prod_{k=1}^{r_i} \theta_{ijk}^{n_{ijk}}, \tag{13.14}$$

where q_i is the number of combinations of the parent nodes pa_i^j, 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 data set D is complete, MLE method can be described as a constrained optimization problem,

$$\max L_D(\theta),$$
$$st. g_{ij}(\theta) = \sum_{k=1}^{r_i} \theta_{ijk} - 1 = 0, \forall i = 1, \ldots, n, \forall j = 1, \ldots, q_i. \tag{13.15}$$

Its global optimum solution is

$$\theta_{ijk} = \frac{n_{ijk}}{n_{ij}}, \tag{13.16}$$

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

13.2   BCN-Based Fault Detection and Inference

The complete monitoring model is established via combining the multivariate causal structure and the Bayesian parameters learning. The qualitative and quantitative relationships among the process variables are revealed to the greatest extent. Then this model is forward used to accurately predict the operation status and detect faults of the critical process variables (i.e., forward inference). Similarly, it also can be inversely used to find the source of the faults (i.e., backward inference). The overall block diagram of the proposed method is shown in Fig. 13.3.

Causality network prediction or inference is to calculate the probability of the hypothesis variables at certain status according to the network topology and con-

Fig. 13.3 Overall design
block diagram
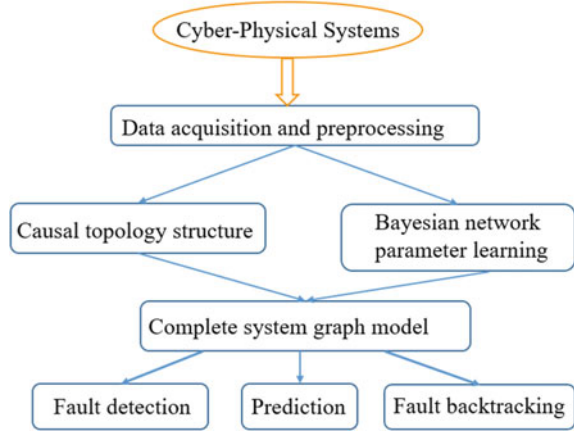
ditional probability distribution of the evidence variable. An inference or query $P(Q = q|E = e_0)$ is to calculate the posterior probability of a query variable Q being at its specific value q in the condition of given evidence e_0 for node E.

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 the evidence in BN to simplify the updating task. JT implements the inference procedure in four steps (Borsotto et al. 2006),

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 separate set S between the two cliques. Every conditional probability distribution of the original BN 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.

$$\phi_X = \prod_{\phi \in \phi_X} \phi. \tag{13.17}$$

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 (architecture is proposed

by Jensen et al. 1990). The projection procedure saves the current potential and assigns a new one to S:

$$\phi_S^{old} \leftarrow \phi_S, \text{ and } \phi_S \leftarrow \sum_{X \setminus S} \phi_X. \tag{13.18}$$

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

$$\phi_Y \leftarrow \phi_Y \frac{\phi_S}{\phi_S^{old}}, \tag{13.19}$$

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,

$$P(Q = q | E = e_0) = \frac{P(Q = q, E = e_0)}{\sum_Q P(Q, E = e_0)}. \tag{13.20}$$

The fault of operational variables is an intervention that has various effects on the production process. The main task in fault detection is to predict the system output and detect whether a fault occurs. The object of causal inference is to find the real root cause under the faulty intervention.

13.3   Case Study

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

13.3.1   Public Data Sets Experiment

Four published data sets proposed by Mooij and Janzing (Leoand et al. 2001) are 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/, which is considered as the benchmark for testing causal detection algorithms. The four data sets are (1) the ground altitude and temperature sampled at 349 stations, US; (2) census income data set which contains weighted census data extracted from the 1994 and 1995 current population surveys conducted by the U.S. Census Bureau.
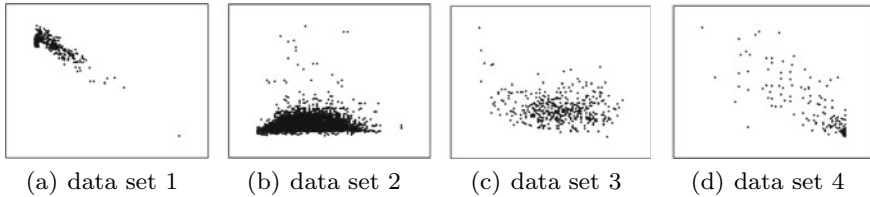
(a) data set 1   (b) data set 2   (c) data set 3   (d) data set 4

Fig. 13.4 Scatter plots of four data sets, **a–d** corresponding to data sets (1)–(4), respectively

Table 13.1 Independence test statistics under different assumption of causal directions

Causal assumption	$x \to y$	$y \to x$
#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 13.2 Causal results of the public data sets

Data sets	#1	#2	#3	#4
Data information	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

The variables include age and wage per hour; (3) the attribute information (age and heart rate) from Cardiac Arrhythmia database; (4) the population with sustainable access to improved drinking water sources (%) total, and the infant mortality rate (per 1000 live births) both sexes, 2006. Each data set consists of two random variables which their cause-effect relationship is known. The four data sets have different attributes, which is sufficient to show the general and comprehensive nature of the data.

Figure 13.4 gives the scatter plots of the selected data sets (1)–(4). Table 13.1 shows the statistics of independence test on x and y for data sets (1)–(4) under different assumption of causal directions. The statistics are calculated separately based on these different assumptions. Comparing the test statistics under two different assumption in Table 13.1, the causal direction of each set all are determined as $x \to y$. Table 13.2 summarizes the causal results obtained by the multivariate causality model. It is found that the test results are consistent to the real causal relationship. We can conclude that the proposed method can correctly identify the causal direction regardless the diversity of data.

Fig. 13.5 The network Bnet0 from the mechanism analysis
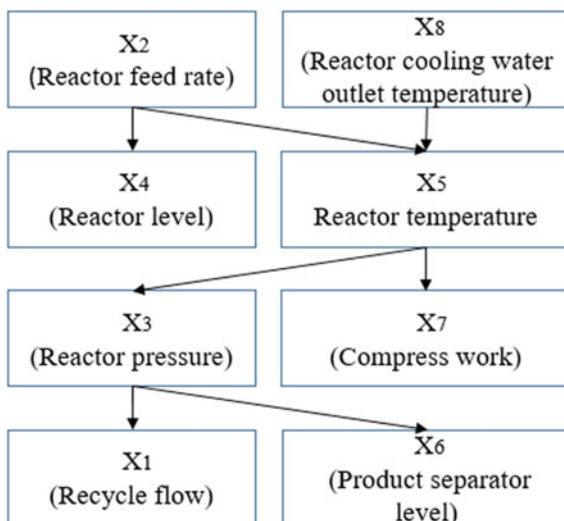
13.3.2   TE Process Experiment

In order to illustrate the applicability of the proposed method in the actual complex industrial process, the network topology of TE process is established and used to predict the alarm variables. TE platform simulates an actual chemical process, a detailed description of the TE process is given in Chap. 4.

Experiment 1: Build Causal Structure

In this experiment, eight important process variables are selected to calculate their causality in order to facilitate the result visualization. From the mechanism analysis of TE process, it is known that when the reactor 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 reactor feed X_2 are the main factors of affecting the reactor temperature X_5. The reactor pressure X_3 changes synchronized with the reactor temperature X_5 according to the general physical principle. In addition, once the chemical 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 also has an obvious influence on the recovered flow X_1 and the material level X_6 in the separator. Now the initial structure of the causality network is built based on the mechanism analysis (including the expert prior knowledge and the intuitive correlation analysis of process variable), named as Bnet0 shown in Fig. 13.5.

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 h to ensure that the data is sufficient to reflect the system process. 500 sampling data are obtained after the equal time decimating. The causal direction of the paired variables is shown in Table 13.3. Three different causality

Table 13.3 Causal direction of TE variables

Variables information	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	$3.4 \times 10^{-4}/8.5 \times 10^{-4}$	$X_2 \rightarrow X_4$
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	$6.6 \times 10^{-6}/2.7 \times 10^{-6}$	$X_3 \rightarrow X_1$

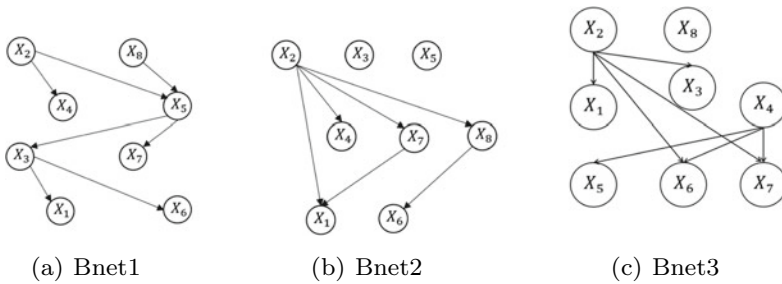(a) Bnet1   (b) Bnet2   (c) Bnet3

Fig. 13.6 The network compare: **a** Bnet1, **b** Bnet2, **c** Bnet3

models are compared, including (1) Bnet1, the proposed multivariate post-nonlinear acyclic causal model, shown in Fig. 13.6a; (2) Bnet2, an alternative network obtained from the traditional BN structure learning method-K2 algorithm which needs to set the node order, shown in Fig. 13.6b; (3) Bnet3, the network structure learned with the expectation maximization (EM) algorithm, shown in Fig. 13.6c.

Comparing the process analysis structure Bnet0 and Bnet1 determined by the proposed Bayesian Causal Network, it is found 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 are not consistent with the mechanism. They show a big gap from the actual physical relationship. It demonstrates that the general BN learning method fails when it is applied to the complex nonlinear systems, while the proposed multivariate causality model proves its superiority.

Table 13.4 Threshold setting for alarm status in different variables

Alarm status	X_1 (km^3/h)	X_2 (km^3/h)	X_3 (kPa)	X_4 (%)	X_5 (°C)	X_6 (%)	X_7 (KW)	X_8 (°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

Experiment 2: Parameter Learning Once the TE network structure is determined, the alarm prediction model can be obtained by parameter learning of this causality structure network. 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 discretize the continuous variables into five-alarm levels by setting different thresholds, shown in Table 13.4.

Here the MLE algorithm is adopted 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 the conditional probability values for all variables are calculated based on the BN 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 Fig. 13.7. Hot plot is used to show the probability since the precise value has nothing meaning for the alarm prediction and inference. The color represents the probability range between 0 and 1.

It should be concerned with the probability value of close to 1. These are the key points 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 13.7a shows the probability of X_5 under the combined action of X_2 and X_8. The abscissa is the status condition of X_8 and X_2, and the ordinate is the probability value for five-alarm status of X_5 displayed in corresponding color. $P(X_5 = 1|X_8 = 1, 2$ and $X_2 = 1) \approx 1$ in the lower left corner of Fig. 13.7a. It means that X_5 occurs the low-low alarm with the probability close to 1 when X_2 and X_8 are in the low-low alarm status. $P(X_5 = 5|X_8 = 4, 5$ and $X_2 = 5) \approx 1$ in the upper right corner of Fig. 13.7a. It means that X_5 occurs the high-high alarm with the probability close to 1 when X_2 and X_8 are in the high-high alarm status. These inference results are consistent with the actual mechanism.

Figure 13.7b–e reflects the probability relationship between bivariate variables. Figure 13.7b shows the probability of X_4 under the action of X_3. $P(X_4 = 5|X_3 =$
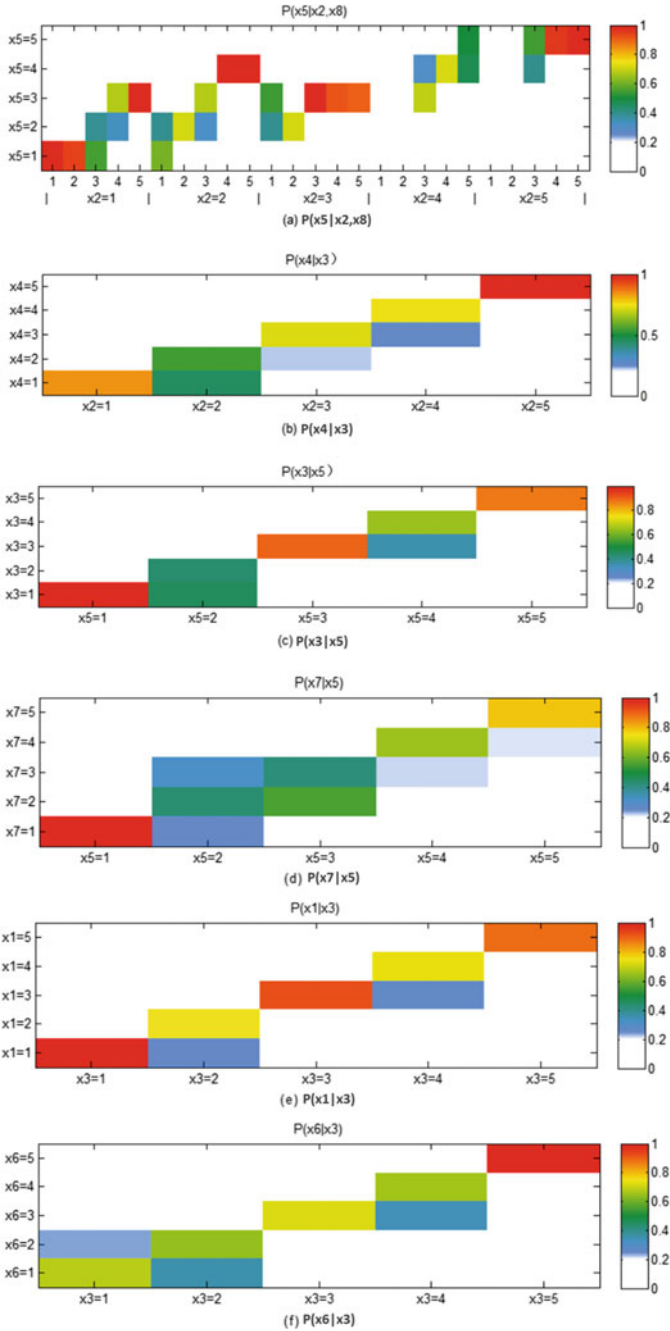
Fig. 13.7 Conditional probability of the descendant variables: **a** $P(X_5|X_8, X_2)$, **b** $P(X_4|X_3)$, **c** $P(X_3|X_5)$, **d** $P(X_7|X_5)$, **e** $P(X_1|X_3)$, **f** $P(X_6|X_3)$

Table 13.5 Alarm level prediction of compress work X_7

No.	X_2	X_8	X_5	X_7	$\hat{X}_7$	Max Prob.
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

5) ≈ 1 in the upper right corner. It means that the probability of X_4 occurs the high-high alarm with the probability close to 1 when X_3 in the high-high alarm status. However, $P(X_4 = 1 | X_3 = 5) = 0$ in the lower right corner. It means that X_4 occurs the high-high alarm with the probability close to 0 when X_3 in the low-low alarm statue. $P(X_4 = 1 \text{ and } X_4 = 2 | X_3 = 2) \approx 0.5$ in the green area. It means the probability of X_4 occurs the low alarm or low-low alarm almost same when X_3 in the low alarm status. Similarly, the inference results obtained from Fig. 13.7c–e are consistent with the mechanism.

Experiment 2: 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 13.5, where $\hat{X}_7$ is the prediction value of X_7.

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 mis-predictions confuse between 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.

13.4   Conclusions

This chapter proposes a multivariate causality model to analyze the causal direction of multivariable and final determine the network topology. The proposed method can describe the system structure more accurate than the traditional BN structure learning method especially when the industrial process is high complex. Combined with the network parameters learning and evidence inference technique, an accurate monitoring and alarm prediction can be performed. The validity of the proposed method is verified via the public data set and TE process. An compact network structure and confident alarm prediction are obtained for the TE process based on the causal analysis and probability inference. Both the methodology and the simulation results show that the proposed multivariate causality model has 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 causal analysis problems in the real world. Developing the efficient algorithm to find the causal relationship of multiple variables based on the general functional causal models is still an important topic. To improve the computational efficiency, a feasible solution is to limit the complexity of the causal structure, such as decreasing the number of direct causes of each variable. Moreover, a smart optimization procedure instead of the exhaustive search should be considered further.

References

Borsotto M, Zhang W, Kapanci E, Pfeffer A, Crick C (2006) A junction tree propagation algorithm for bayesian networks with second-order uncertainties. In: 18th IEEE international conference on tools with artificial intelligence, 2006. ICTAI '06

Chen X, Wang J, Zhou J (2018) Fault detection and backtrace based on graphical probability model. In: 2018 prognostics and system health management conference (PHM-Chongqing)

Giga M (2014) Statistical tests, test of independence. Nihon Ika Daigakkai Zasshi 10(2):115–119

Hipel KW, Kilgour DM, Fang L (2011) The graph model for conflict resolution. Wiley encyclopedia of operations research and management science

Hong Y, Hao Z, Mai G, Chen B, Rui X (2017) Inferring causal direction from multi - dimensional causal networks for assessing harmful factors in security analysis. IEEE Access 5:20009–20019

Hyvärinen A, Zhang K, Shimizu S, Hoyer PO (2010) Estimation of a structural vector autoregression model using non-gaussianity. J Mach Learn Res 11(2010):1709–1731

Ishak MB, Leray P, Amor NB (2011) A two-way approach for probabilistic graphical models structure learning and ontology enrichment. In: International conference on knowledge engineering ontology development

Jensen FV, Lauritzen SL, Olesen KG (1990) Bayesian updating in causal probabilistic networks by local computations. Comput Stat Quaterly 4:269–282

Jiang Y, Deng Z, Chung FL, Wang S (2015) Multi-task TSK fuzzy system modeling using inter-task correlation information. Inf Sci 298:512–533

Johnson RA, Bhattacharyya GK (2016) Statistics and causality: methods for applied empirical research. Wiley, Hoboken

Leoand M, Russell E, Braatz R (2001) Tennessee eastman process. Springer, London

Li X, Zhao L, Wei L, Yang MH, Fei W, Zhuang Y, Ling H, Wang J (2016) Deepsaliency: multi-task deep neural network model for salient object detection. IEEE Trans Image Process Publ IEEE Signal Process Soc 25(8):3919

Pearl J (1986) Fusion, propagation, and structuring in belief networks. Artif Intell 29(3):241–288

Shimizu S, Hoyer PO, Kerminen A (2006) A linear non-gaussian acyclic model for causal discovery. J Mach Learn Res 7(4):2003–2030

Shimizu S, Inazumi T, Sogawa Y, Hyvarinen A, Kawahara Y, Washio T, Hoyer PO, Bollen K (2011) DirectLiNGAM: a direct method for learning a linear non-gaussian structural equation model. J Mach Learn Res 12(2):1225–1248

Zhu J, Ge Z, Song Z, Zhou L, Chen G (2017) Large-scale plant-wide process modeling and hierarchical monitoring: a distributed bayesian network approach. J Process Control, pp 91–106