Received February 7, 2020, accepted February 20, 2020, date of publication February 27, 2020, date of current version March 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976808

Graph-Based Method for Fault Detection in the Iron-Making Process

RUQIAO AN[1], CHUNJIE YANG[1], (Senior Member, IEEE), AND YIJUN PAN[2,3,4]

[1]Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
[2]Key Laboratory of Networked Control Systems, Chinese Academy of Sciences, Shenyang 110016, China
[3]Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[4]Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China

Corresponding author: Chunjie Yang (cjyang@iipc.zju.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61933015.

ABSTRACT Since the iron-making process is performed in complicated environments and controlled by operators, observation labeling is difficult and time-consuming. Therefore, unsupervised fault detection methods are a promising research topic. Recently, an unsupervised graph-based change point detection method has been introduced, and the graph of observations is constructed by the minimum spanning tree. In this paper, a novel fault detection method based on the graph for an iron-making process is proposed, and a weight calculation method for constructing the minimum spanning tree is introduced. The Euclidean distance and Mahalanobis distance are combined to calculate the weights in the minimum spanning tree, which contain important relations of variables. The distance calculation method is determined by the correlation coefficients of variables. Each testing observation is set as a change point candidate, and a change point candidate divides the observations into two groups. The number of a special type of edge in the minimum spanning tree is used as a fault detection statistic. That special edge connects two observations from two different groups. The minimum number of that type of edge corresponding to the change point candidate is a true change point. Finally, numerical simulation is used to test the power of the proposed method, and a real iron-making process including low stock, cooling, and slip faults is implemented to illustrate the effectiveness of fault detection in industrial processes.

INDEX TERMS Fault detection, graph, iron-making process, Mahalanobis distance, minimum spanning tree.

I. INTRODUCTION

In recent decades, industrial processes have become increasingly complicated, and a large quantity of data has been collected. Since industrial processes are the foundation of national development, ensuring safe and reliable operation is important [1]. Due to the complicated production principles of industrial processes, observation labeling is difficult. Moreover, sensor errors can occur, and the use of experts for observation labeling is time-consuming and expensive. Therefore, unsupervised process monitoring methods must be researched for application in industrial process.

Iron-making and steel-making, which have been widely researched recently, are important parts of modern industrial processes [2]–[5]. In iron-making and steel-making, the blast furnace, which consumes more than 70% of the total energy, is a key unit. The blast furnace is a metallurgical furnace used in the manufacture of the industrial metals, and it works under high temperature and high pressure [6], [7]. The iron-making process consists of a blast furnace and five pieces of smelting auxiliary system equipment. In the iron-making process, solid raw materials including coke and ore are placed in the top of the blast furnace in specific amounts and proportions, and hot wind is blown from the furnace bosh [8].

In the iron-making process, there are some typical faults including low stock line, cooling, and slip. To ensure the safety and reliability of the iron-making process, these faults should be addressed [9], [10]. The most widely used fault detection methods are data-based methods, and the expert system is auxiliary in the iron-making process [11], [12]. The expert knowledge and experience that accumulate during production are difficult to describe by exact mathematical language. Moreover, it is necessary to periodically update the expert system according to the actual production process

The associate editor coordinating the review of this manuscript and approving it for publication was Mervat Adib Bamiah.

and the equipment operational condition to maintain validity. However, due to the low grade of raw materials, the characteristics are constantly changing. The real-time update speed of the expert system may not satisfy the time-varying system of the iron-making process. Moreover, since the iron-making process is operated by experienced operators, parameter adjustments and production status identifications vary. Under some conditions, operators decide that the minor changes do not need to be controlled, which may affect the observations. It is difficult to obtain observations without disturbances. The model constructed using corrupted observations will lead to poor fault detection results. Therefore, unsupervised data-based process monitoring methods are necessary in the iron-making process.

Unsupervised process monitoring methods have been researched in recent decades. Jove et al. described a novel approach using a visual tool for the detection of faults in industrial processes via unsupervised and projectionist techniques [13]. Escobar et al. proposed a combined generative topographic mapping and graph theory approach for unsupervised nonlinear data visualization and fault identification [14]. An et al. adopted an unsupervised graph method for fault detection, and the time interval was used to improve the power [15]. A support vector clustering-based probabilistic approach was developed for unsupervised chemical process monitoring and fault classification by Yu [16]. Spyridon et al. proposed a fault detection scheme based on the unsupervised training of a generative adversarial network [17]. Bezerra et al. proposed applying typicality and eccentricity data analytics, a fully autonomous algorithm, to address the problem of fault detection in industrial processes [18].

Recently, Chen et al. proposed an unsupervised graph-based change point detection method [19], [20]. Compared with the above unsupervised process monitoring methods in the industrial processes, the graph-based method can be used to handle various types of data, such as non-Gaussian data and nonlinear data, and accords to the characteristics of observations collected from iron-making processes. Moreover, the graph-based method applies a specific analytic formula to build a change point detection statistic, which is simple to calculate. The purpose of the graph-based method is to determine whether the change point occurs in the time series. The goal of fault detection in iron-making processes is to find the abnormal events from the collected data matrix, which can be regarded as the change points in the observations. Therefore, in this paper, the graph-based change point detection method is used for fault detection in the iron-making processes.

There are three steps for detecting change points from observations based on the graph method. First, the graph is constructed. Each observation collected from the iron-making process is treated as a node in the graph. There are many ways to build the graph, such as minimum spanning tree, minimum distance pairing and nearest neighbor graph. Second, the statistic is calculated. Each observation is regarded as a change point candidate, and the observations are divided into two groups. The statistic is the count of the number of a special type of edge in the graph. The special edge links two observations from different groups. Finally, the change point is detected. The change point candidate corresponding to the minimum statistic should be a true change point. Chen et al. used the graph-based change point detection method to judge the writing style of an author [19]. To the best of our knowledge, the graph-based change point detection method has barely been applied for fault detection in the iron-making process. Spectral graph analysis theory was used for process monitoring, and it applies different principles for constructing the model [21].

An et al. used the time interval to improve the process monitoring power based on the graph method. However, the relations between variables are not considered, which may be the reason why the method is ineffective for some faults [15]. The graph-based change point detection method uses the Euclidean minimum spanning tree to build the graph. In the iron-making process, some variables are used to describe the production status. The collected variables are related, and the Euclidean minimum spanning tree ignores the relationships between variables. Therefore, the Mahalanobis distance is considered for calculating the weights in the minimum spanning tree in this paper. The Mahalanobis distance contains the relations between variables but could increase the influence of minor changes in variables [22].

Based on the graph method, a minimum spanning tree is constructed. In this paper, a novel weight calculation method using the minimum spanning tree is proposed. The Euclidean distance and Mahalanobis distance are both used to calculate the distance between observations. Some variables are calculated using the Euclidean distance, while other variables are calculated using the Mahalanobis distance. The distance calculation method is determined by the correlation coefficients of variables, and the parameters λ, β are introduced. If the correlation coefficients of one variable with the others are all less than a threshold, that variable will be considered to be less correlated with other variables, and the Euclidean distance will be calculated. The sum of the distances between the variables is regarded as the weight for constructing the minimum spanning tree. The other steps for change point detection based on the proposed method are the same as those of the origin graph-based method.

This paper is organized as follows. In section 2, the problem formulation is introduced. The graph-based change point detection method and the proposed method for fault detection in the iron-making process are presented in section 3. In section 4, the simulation results of a numerical simulation and a real iron-making process are illustrated. Finally, the conclusions are summarized.

II. PROBLEM FORMULATION

In the iron-making process, the observation x_i with m variables is collected by sensors. The purpose of the paper is to verify whether the fault occurs after t observations. In this paper, only the situation in which a single change point occurs is discussed for convenience. If there are multiple change points in the process, the data can be divided into parts to detect the faults. Under normal conditions, the processes are stable, and the observations should follow the same distribution. Therefore, the problem can be defined by the following hypotheses based on the distribution of the observations in this paper. The null hypothesis means that there are no faults in the testing observation, and all the observations follow the same distribution, which is shown as (1).

$$H_0 : x_i \sim F_0, \quad i = 1, 2, \ldots, n \tag{1}$$

where n is the total number of observations, and F_0 is the data distribution. In contrast, the alternative hypothesis is shown as (2).

$$H_1 : x_i \sim \begin{cases} F_0, & i = 1, 2, \ldots \tau \\ F_1, & i = \tau + 1, \tau + 2, \ldots, n \end{cases} \tag{2}$$

where τ is a change point. The distributions F_0 and F_1 are derived from the data.

In this paper, two assumptions need to be discussed.

Assumption 1: The collected observations in the iron-making process are independent.

Assumption 2: The influence of noise is ignored in this paper for convenience.

The variables in the iron-making process are correlated to improve the reliability. The observations can be made independent by extending the sampling time. The graph-based method compares the data distributions to detect the faults. Noise exists throughout the iron-making process, and the faults do not influence the distribution of the noise. Therefore, assumption 2 is acceptable. More details can be found in the literature [15].

III. UNSUPERVISED WEIGHT GRAPH CHANGE POINT DETECTION METHOD

A. GRAPH-BASED CHANGE POINT DETECTION METHOD

Suppose that there is a data matrix $X \in R^{n\times m}$ containing n observations with m variables. The purpose of the graph method is to find the change point τ from n observations. The null hypothesis H_0 and the alternative hypothesis H_1 are the same as those in the above section.

As mentioned, there are three steps for detecting the change points by the graph method. Before computing the statistics, the graph is constructed by two steps. The first step is to calculate the Euclidean distances between two observations. Let $D \in R^{n\times n}$ be a matrix such that d_{ij} is the Euclidean distance between observations i^{th} and j^{th}, $i, j = 1, \ldots, n$. D is the weighted adjacency matrix of a complete edge-weighted simple graph, $G = (V, E)$, where $i \in V$ is an observation, and the weight of edge $(i, j) \in E$ is the Euclidean distance between observations i^{th} and j^{th}.

The second step is to construct the minimum spanning tree based on the weight matrix D. Each observation is set as a change point candidate, and each change point candidate divides the observations into two groups by the change point candidate. The observations are divided into those before the change point candidate and those after. In the graph, if two linked observations are obtained from different groups, then the edges are recorded.

In other words, suppose that there are n observations; each observation is regarded as a change point candidate in order. The observations are divided into two groups n times based on each change point candidate. The minimum spanning tree reflects the relations between observations, and the edges in the minimum spanning tree are used to build the fault detection statistic. For example, there is a true change point at the 20^{th} sampling point, and the observations follow different distributions before and after the 20^{th} sampling point. When the 20^{th} observation is set as a change point candidate, an observation obtained from the first 20 observations prefers to connect to the one derived from the first 20 observations. When the count of the recorded edges is low, the null hypothesis is rejected. The description can be formulated as (3):

$$\begin{aligned} R_G(t) &= \sum_{(i,j)\in G} I_{g_i(t)\neq g_j(t)} \\ g_i(t) &= I_{i>t} \end{aligned} \tag{3}$$

where $R_G(t)$ is the number of the edges in the tree connecting observations from different groups. G is the minimum spanning tree of observations, $t = 1, 2, \ldots, n$ is a change point candidate and I_x is an indicator function shown in (4) [23].

$$I_x = \begin{cases} 1, & \text{if } x \text{ is true} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The third step is to calculate the statistics and to detect the change point. The change point candidate corresponding to the minimum number of edges is a true change point.

B. IMPROVED GRAPH-BASED CHANGE POINT DETECTION METHOD

Since production is complicated, the collected observations may be corrupted during the iron-making process [24]. In this paper, an unsupervised graph-based change point detection method is adopted to realize fault detection in an iron-making process, and a novel weight calculation method is proposed. The minimum spanning tree can be constructed by calculating the Euclidean distances between two observations. However, the variables in an iron-making process are dependent to improve reliability [25]. Compared with the Euclidean distance, the Mahalanobis distance contains the relationships between variables; it is suitable for calculating the weights of the edges in an iron-making process but would magnify the influence of minor changes in variables.

In this paper, considering the advantages of the Euclidean distance and the Mahalanobis distance, the two distances are combined and used to calculate the weights in the minimum spanning tree. The correlation coefficients of the variables are calculated first. If the correlation coefficients of one variable with the others are all less than 0.6, then this variable is less related to all the other variables. The Euclidean distances of this variable will be calculated for all observations. The corresponding parameter λ is set to 1; otherwise, the variables will be calculated using the Mahalanobis distance, and the corresponding parameter β is set to 1. Equation (5) presents the formula to compute the distance for the improved graph-based model.

$$w_{ij} = \lambda_{ij}\|x_i - x_j\|_2 + \beta_{ij}\sqrt{(x_i - x_j)S^{-1}(x_i - x_j)'} \tag{5}$$

where x_i, x_j are the i^{th}, j^{th} observations, $\lambda_{ij} \in \{0, 1\}$ and $\beta_{ij} \in \{0, 1\}$ are two parameters, which could be regarded as the correlation factor, and are used for choosing the distance calculation method. w_{ij} is the weight of the edge connecting two observation x_i, x_j for constructing the minimum spanning tree. S is the covariance matrix of some observations without change points, which could be obtained by the original graph-based change point detection method.

If the correlation coefficient of two variables is less than 0.6, then these two variables can be regarded to be less correlated. Thus, the threshold is set to 0.6. The correlation coefficients are used to obtain variables that are less correlated, which do not need to calculate the Mahalanobis distance. A suitable threshold can make good use of the Mahalanobis distance and Euclidean distance and obtain better fault detection performance. The fault detection statistic is the number of edges. An edge connects two observations from two different groups, and the groups are divided by a change point candidate. A regular number of edges corresponds to the occurrence of a fault.

C. FAULT DETECTION BASED ON AN IMPROVED GRAPH METHOD

The observations $x_i, i = 1, 2, \ldots, n$ with m variables are collected from an iron-making process. Since the covariance matrix of some observations without change points needs to be calculated for the Mahalanobis distance, the graph-based change point detection method introduced in the above section is used to obtain observations with the same distribution, defined by $y_i, i = 1, 2, \ldots, 10$. The purpose of this paper is to determine whether there is a change point after τ observations. Fault detection based on the improved graph method can be summarized as follows.

Step 1: Correlation coefficient calculation

The observations $y_i, i = 1, 2, \ldots, 10$ are used to calculate the correlation coefficients between variables. The correlation coefficient calculation formula is shown as (6) [26]:

$$r_{ij} = \frac{cov(m_i, m_j)}{\sqrt{var(m_i)var(m_j)}} \tag{6}$$

where m_i, m_j are the i^{th}, j^{th} variables, $cov(m_i, m_j)$ is the covariance of m_i, m_j, and $var(m_i)$, $var(m_j)$ are the variances of m_i, m_j.

Step 2: Variable choice

In this paper, the distances between observations are calculated by the Mahalanobis distance and Euclidean distance. If the correlation coefficients of one variable with the other variables are all less than 0.6, then the variable will be regarded as independent. These variables are calculated using the Euclidean distance as the weights of the edges. The corresponding parameter λ is set to be 1, which could be formulated by $\lambda_{(i,:)} = \lambda_{(:,i)} = 1, \beta_{(i,:)} = \beta_{(:,i)} = 0$; otherwise, the variables will be calculated the Mahalanobis distance, and the corresponding parameter β is set to be 1, $\lambda_{(i,:)} = \lambda_{(:,i)} = 0$, $\beta_{(i,:)} = \beta_{(:,i)} = 1$.

Step 3: Euclidean distance calculation

Calculate the Euclidean distances between observations shown in (7), collected by matrix A_1:

$$a_{1ij} = \|x_i - x_j\|_2 \tag{7}$$

where x_i, x_j are the i^{th}, j^{th} observations [27].

Step 4: Mahalanobis distance calculation

Calculate the Mahalanobis distances between observations shown in (8), collected by matrix A_2:

$$a_{2ij} = \sqrt{(x_i - x_j)S^{-1}(x_i - x_j)'} \tag{8}$$

where x_i, x_j are the i^{th}, j^{th} observations, and S is the covariance matrix of the normal observations [28].

Step 5: Weight matrix calculation

Calculate the weights based on the Euclidean distance and the Mahalanobis distance shown in (9), collected by matrix B:

$$B = \lambda. * A_1 + \beta. * A_2 \tag{9}$$

where λ, β is the matrix calculated in Step 2, and A_1, A_2 corresponds to the distance matrix.

Step 6: Graph construction

Construction of the minimum spanning tree of the graph represented by the weighted adjacency matrix B [29].

Step 7: Statistic calculation

The number of the edges connecting two observations obtained from different groups is regarded as the fault detection statistic, and the groups are divided by a change point candidate. The number of edges $R_G(t)$ can be formulated as (10).

$$\begin{aligned} R_G(t) &= \sum_{(i,j)\in G} I_{g_i(t)\neq g_j(t)} \\ g_i(t) &= I_{i>t} \end{aligned} \tag{10}$$

Since the method of counting edges is related to the position of the change point candidate, $R_G(t)$ is normalized to improve the interpretability. $Z_G(t)$ is normalized by using the mean and variance of $R_G(t)$ shown in (11). Since a lower value of $R_G(t)$ means that a fault occurred, the minus is used for convenience. A larger value of $Z_G(t)$ means that a fault occurred.

$$Z_G(t) = -\frac{R_G(t) - E[R_G(t)]}{\sqrt{Var[R_G(t)]}} \tag{11}$$

where

$$\begin{aligned} E[R_G(t)] &= p_1(t)|G| \\ Var[R_G(t)] &= p_2(t)|G| + (\frac{1}{2}p_1(t) - p_2(t))\sum_i |G_i|^2 \\ &\quad +(p_2(t) - p_1^2(t))|G|^2 \end{aligned}$$

TABLE 1. The correlation coefficient of 6 variables.

Variable	1	2	3	4	5	6
1	1	-0.1410	-0.1085	-0.0489	0.8222	-0.6057
2	-0.1410	1	-0.0107	-0.7807	0.1331	0.1937
3	-0.1085	-0.0107	1	-0.0063	-0.3500	0.0705
4	-0.0489	-0.7807	-0.0063	1	0.0111	-0.0973
5	0.8222	0.1331	-0.3500	0.0111	1	-0.4676
6	-0.6057	0.1937	0.0705	-0.0973	-0.4676	1

$$p_1(t) = \frac{2t(n-t)}{n(n-1)}$$
$$p_2(t) = \frac{4t(t-1)(n-t)(n-t-1)}{n(n-1)(n-2)(n-3)} \tag{12}$$

where $|G|$ is the number of the edges in graph G, G_i is a subgraph containing all edges connected to observation x_i, $|G_i|$ is the number of edges in G_i, $E[R_G(t)]$ is the expectation of $R_G(t)$, and $Var[R_G(t)]$ is the variance of $R_G(t)$.

Step 8: Fault detection

Under normal conditions, the observations follow the same distribution, and the edges in the minimum spanning tree should be irregular. The irregular number of the edge means that there is no element greater than others obviously. If there is a change point and the testing observations follow different distributions, then the count of the recorded edges is low, the null hypothesis is rejected. Therefore, under normal conditions, the elements in vector $Z_G(t)$ should be irregular. When one element is greater than the other elements, a fault occurred.

In conclusion, the steps for fault detection based on the proposed method are outlined, and a simple Gaussian example is used to explain the proposed fault detection method. The testing data contain 40 observations and follow the distributions $N \sim (0, I_6), N \sim ((5, 5, 5, 5, 5, 5)', I_6)$. There is a change point at the 20^{th} sampling point. The correlation coefficients of 6 variables are listed in Table 1.

From Table 1, it can be seen that the correlation coefficients of variable 3 with the other 5 variables are all less than 0.6. Therefore, variable 3 is less correlated with other variables, and the Euclidean distance of variable 3 is used to calculate the weights of the edges between the node which corresponds to observation 3 and the other nodes. The other 5 variables are calculated using the Mahalanobis distance for 40 testing observations as the weights. The sum of the distances is used as a final weight for constructing the minimum spanning tree, and the connection order of the testing observations is presented in Table 2. The numbers of edges connecting two observations derived from different groups are listed in Table 3. t is the change point candidate, and $R_G(t)$ is the number of edges. The simulation results are shown in Fig. 1.

The connection orders of 40 observations in the minimum spanning tree are shown in Table 2. Each observation is set as a change point candidate, and the edges connecting two observations derived from different groups are counted in Table 3. For example, the 20^{th} sampling point is set as a change point candidate, and there is only the 19^{th} observation connects to the 26^{th} observation. Thus, the number of the

TABLE 2. The connection order of the testing observations in the minimum spanning tree.

No.	Connection Order	No.	Connection Order
1	$1 \rightarrow 12$	21	$26 \rightarrow 22$
2	$12 \rightarrow 7$	22	$22 \rightarrow 31$
3	$12 \rightarrow 16$	23	$31 \rightarrow 21$
4	$16 \rightarrow 3$	24	$21 \rightarrow 29$
5	$3 \rightarrow 4$	25	$29 \rightarrow 40$
6	$16 \rightarrow 5$	26	$40 \rightarrow 33$
7	$5 \rightarrow 20$	27	$40 \rightarrow 25$
8	$20 \rightarrow 6$	28	$25 \rightarrow 34$
9	$6 \rightarrow 18$	29	$25 \rightarrow 38$
10	$18 \rightarrow 15$	30	$38 \rightarrow 24$
11	$6 \rightarrow 13$	31	$24 \rightarrow 36$
12	$1 \rightarrow 14$	32	$36 \rightarrow 32$
13	$7 \rightarrow 11$	33	$36 \rightarrow 35$
14	$11 \rightarrow 2$	34	$32 \rightarrow 30$
15	$2 \rightarrow 17$	35	$36 \rightarrow 37$
16	$13 \rightarrow 10$	36	$37 \rightarrow 39$
17	$10 \rightarrow 9$	37	$30 \rightarrow 28$
18	$9 \rightarrow 8$	38	$28 \rightarrow 23$
19	$17 \rightarrow 19$	39	$28 \rightarrow 27$
20	$19 \rightarrow 26$		

TABLE 3. The number of edges.

t	$R_G(t)$	t	$R_G(t)$
1	2	21	3
2	4	22	5
3	6	23	6
4	5	24	8
5	7	25	11
6	10	26	9
7	12	27	10
8	13	28	9
9	13	29	9
10	13	30	9
11	11	31	7
12	10	32	7
13	8	33	8
14	7	34	7
15	8	35	8
16	5	36	6
17	5	37	6
18	3	38	4
19	3	39	3
20	1	40	0

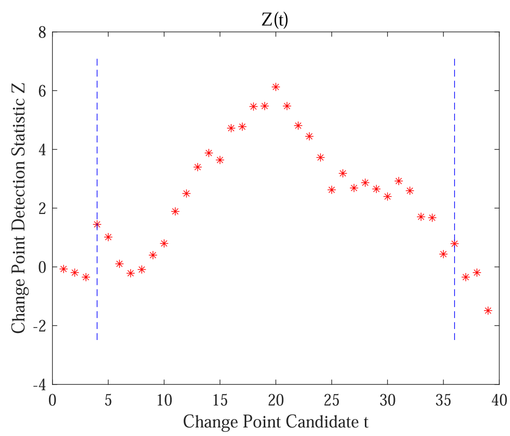


FIGURE 1. A simple Gaussian example for illustrating the proposed fault detection method.

edges is 1, which could be found in Table 3. From Table 3, the number of the edges at the 20^{th} sampling point is 1, which is the minimum. Due to the influence of the change point candidate position, the first 5% and final 5% of sampling points are not considered as change points. The fault detection result based on the proposed method is presented in Fig. 1, which is the normalization of Table 3. The change point detection statistic is maximum at 20^{th} sampling point. There is a true change point at the 20^{th} sampling point, and the change point could be detected exactly based on the proposed method.

IV. SIMULATION

In this section, a numerical simulation is applied for testing the effectiveness of the proposed method, and then the proposed fault detection method is implemented in a real iron-making process.

A. TEST METHODOLOGY

There are some testing observations including a fault, and the proposed method is used on these data. In a numerical simulation, 40 simulated observations are obtained by MATLAB, and there is a change point in the observations at the 20^{th} sampling point. The first 20 observations follow the different distribution from the after 20. The proposed method is used for detecting the different data distributions. The correlation coefficients of the variables are calculated first. The time at which the fault is detected is recorded and compared with the true time of the change point occurrence to verify the power of the proposed method via numerical simulation. In the iron-making process, the correlation coefficients of the variables are calculated first based on observations with the same distribution. The 120 testing observations include three faults, and low stock line, cooling, and slip are used to detect the faults. Each matrix contains 40 observations, and the fault occurs at the 20^{th} sampling point. The time at which the fault is detected is recorded based on the proposed method in the iron-making process. The detect time and the real fault occurrence time are compared to test the power of the proposed method.

B. NUMERICAL SIMULATION

In this part, a numerical simulation model proposed by Acala et al. in 2009 is used to test the fault detection power of the proposed method [30]. The model is constructed as (13):

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} -0.2310 & -0.0816 & -0.2662 \\ -0.3241 & 0.7055 & -0.2158 \\ -0.217 & -0.3056 & -0.5207 \\ -0.4089 & -0.3442 & -0.4501 \\ -0.6408 & 0.3102 & -0.2372 \\ -0.5655 & -0.433 & -0.5938 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \tag{13}$$

where t_1, t_2, and t_3 are zero-mean random variables with standard deviations of 1, 0.8 and 0.6, respectively, i.e., $t_1 \sim N(0, 1)$, $t_2 \sim N(0, 0.8)$, and $t_1 \sim N(0, 0.6)$.

The testing data are generated by the following formula (14):

$$x = x^* + \xi_i f \tag{14}$$

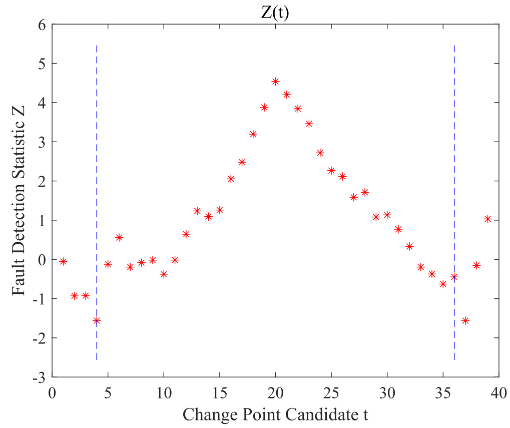


FIGURE 2. Fault detection result based on the graph method in the numerical simulation.

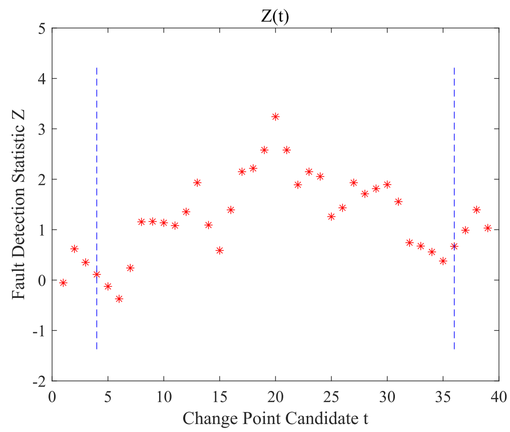


FIGURE 3. Fault detection result based on the proposed method in the numerical simulation.

where x^* is the normal data obtained by formula (13). The direction ξ_i is randomly chosen from six possible variable directions with uniform probability, and f is the fault magnitude following $N \sim (5, 0.1)$. In this section, 40 observations with six variables are used for testing the fault detection power of the proposed method, and there is a change point at the 20^{th} sampling point. The simulation results are presented in Fig. 2 to Fig. 3.

As shown in Fig. 2, the maximum value of $Z_G(t)$ appears at the 20^{th} sampling point, which means that the number of edges connecting two different groups is the minimum. Thus, a change point would exist, and the graph-based method could detect the fault exactly. Fig. 3 is the simulation result for fault detection based on the proposed method. As with the graph-based method, the maximum value of $Z_G(t)$ appears at the 20^{th} sampling point. Moreover, it can be seen that the variance of three larger statistics based on the proposed method is greater than that of the graph method from Fig. 3, which could reduce the false alarm rate. From Fig. 2 to Fig. 3, both methods could detect the change point exactly. Therefore, the proposed method using the Mahalanobis distance to solve the relations between variables is suitable.

TABLE 4. The variables in an iron-making process.

No.	Variable	No.	Variable
1	Permeability Index	10	Hot Air Pressure
2	Standardized Air Speed	11	Actual Air Speed
3	Cold Air Flow	12	Hot Air Temperature
4	Blast Momentum	13	Top Temperature 1
5	Blast Furnace Bosh Gas Volume	14	Top Temperature 2
6	Blast Furnace Bosh Gas Index	15	Top Temperature 3
7	Theoretical Combustion Temperature	16	Top Temperature 4
8	Top Pressure	17	Resistance Coefficient
9	Total Pressure Drop	18	Blast Humidity

TABLE 5. Comparison of two distances in an iron-making process.

Fault	Euclidean Distance	Mixed Distance
Low Stock Line	N/A	25
Cooling	25	20
Slip	20	20

C. FAULT DETECTION IN AN IRON-MAKING PROCESS

In a blast furnace, complex physical and chemical reactions occur at all times. It is crucial to ensure stability during the iron-making process. In general, there are three typical faults in the iron-making process: low stock line, cooling and slip. The raw materials are put in the top of the blast furnace, and a low stock line means the raw materials cannot descend normally. The raw material line is lower than the normal level of 0.5 m for 1 hour. The low stock line disturbs the normal distribution of raw materials. The air permeability of the material column is poor, and the gas flow is chaotic. The raw materials cannot be heated and reduced normally, which is an important cause of cooling and slip faults. The reduction of the wind volume is one of the best methods for handling the low stock line fault [7].

The temperature in a blast furnace should be kept stable; values that are too high or too low are abnormal. There are some reasons for cooling fault including low stock lines, tuyere leakage and failed cooling walls, which could be solved by increasing coal injection. A slip fault occurs when the raw materials fall down suddenly. The poor quality of the raw materials, abnormal airflow distribution and irregular furnace wall are the main reasons for the slip. Slip can be solved by appropriately reducing the wind volume [9], [15].

In this part, the observations are collected from an iron-making company in China, and low stock line, cooling, and slip faults are used to illustrate the power of the proposed method. There are 18 variables, as listed in Table 4. The testing data consist of 40 observations, and the fault occurs at the 20^{th} sampling point. The fault detection results are presented in Fig. 4 to Fig. 9, and the detect times are listed in the Table 5.

Fig. 4 and Fig. 5 present the fault detection results of a low stock line fault based on the graph-based method and the proposed method. The maximum value of $Z_G(t)$ appears at the 18^{th} sampling point based on the graph method, which means the fault occurs at the 18^{th} sampling point. However,

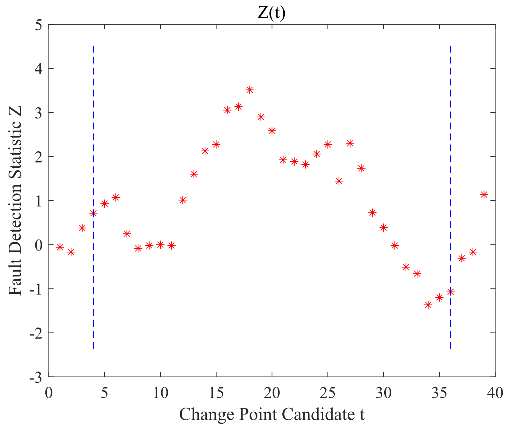


FIGURE 4. The simulation result of low stock line in an iron-making process, based on the graph method.

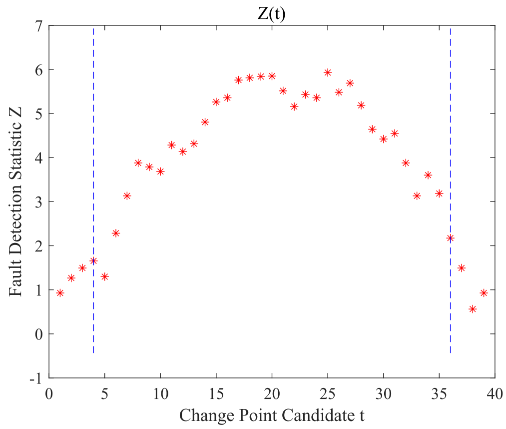


FIGURE 5. The simulation result of low stock line in an iron-making process, based on the proposed method.

the fault occurs at the 20^{th} sampling point, and the graph method shown in Fig. 4 based on the Euclidean minimum spanning tree is invalid. The fault do not occur at 18^{th} sampling point, but is detected. The result in Fig. 4 is invalid in part because the variables have relations in the low stock line fault, and the Euclidean distance does not contain the relations of the variables. As shown in Fig. 5, the maximum value of $Z_G(t)$ appears at the 25^{th} sampling point, and the proposed method could detect the fault at the 25^{th} sampling point. For a low stock line fault, the proposed method is more effective than the graph-based method. For the cooling fault, the maximum value of $Z_G(t)$ occurs at the 25^{th} sampling point shown in Fig. 6, while the maximum value of $Z_G(t)$ appears at the 20^{th} sampling point shown in Fig. 7. Compared with the graph method, the proposed method could detect the fault exactly. For a slip fault, both methods obtain a maximum value of $Z_G(t)$ at the 20^{th} sampling point, and they exhibit good simulation results, as shown in Figs. 8–9. The fault detect time based on the two methods is listed in Table 5. The detect time is the sampling point at which the fault is detected, and the true time of fault occurrence

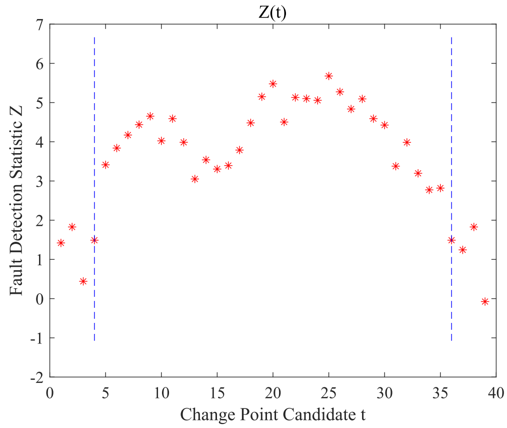


FIGURE 6. The simulation result of cooling in an iron-making process, based on the graph method.

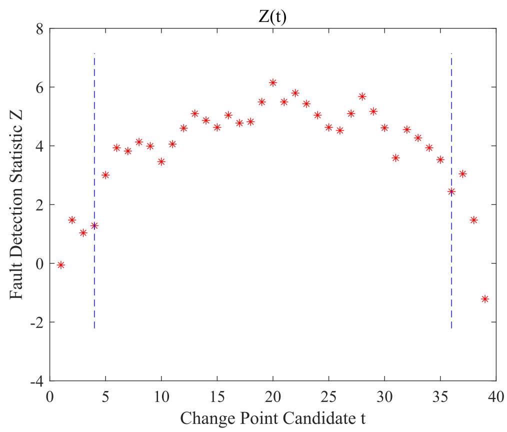


FIGURE 7. The simulation result of cooling in an iron-making process, based on the proposed method.

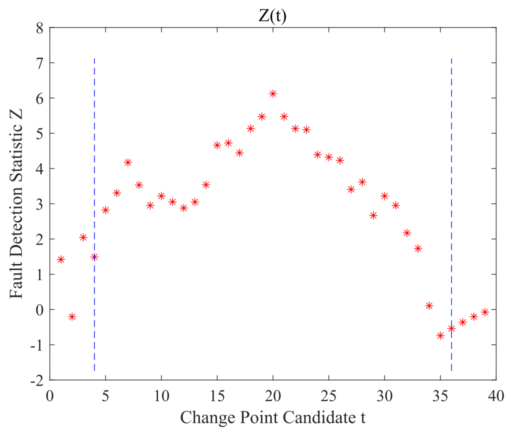


FIGURE 8. The simulation result of slip in an iron-making process, based on the graph method.

is the 20^{th} sampling point. Therefore, the proposed method based on a novel distance calculation method considering the relations of variables is powerful for fault detection in an iron-making process.

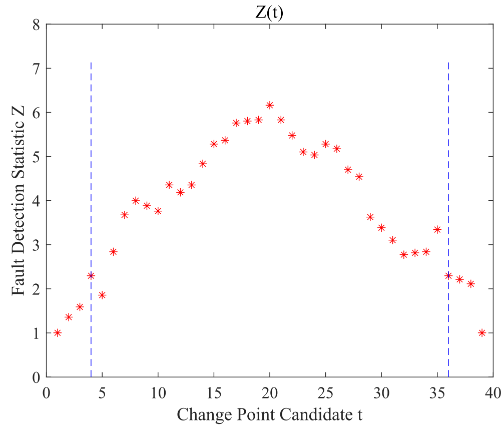


FIGURE 9. The simulation result of slip in an iron-making process, based on the proposed method.

V. CONCLUSION

In this paper, a graph-based fault detection method for an iron-making process is discussed, and a novel distance calculation method in the minimum spanning tree is proposed. The Euclidean distance and the Mahalanobis distance between two observations are calculated, and the parameters are introduced. The parameter is obtained according to the correlation coefficients of variables. The number of edges connecting two observations obtained from two groups is used as a fault detection statistic. The numerical simulation and a practical iron-making process are implemented to test the power of the proposed method. In the future, real-time fault detection based on the graph method should be researched.

REFERENCES

[1] 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, Mar. 2013.
[2] P. Zhou, C. Wang, M. Li, H. Wang, Y. Wu, and T. Chai, "Modeling error PDF optimization based wavelet neural network modeling of dynamic system and its application in blast furnace ironmaking," *Neurocomputing*, vol. 285, pp. 167–175, Apr. 2018.
[3] J. L. Provis and S. A. Bernal, "Geopolymers and related alkali-activated materials," *Annu. Rev. Mater. Res.*, vol. 44, no. 1, pp. 299–327, Jul. 2014.
[4] A. Agrawal, S. C. Kor, A. R. Choudhary, M. K. Agarwal, S. Kundu, and V. R. Tripathi, "A novel method to detect the water leakage from tuyere nose cooling circuit in blast furnace," *Ironmaking Steelmaking*, vol. 43, no. 10, pp. 744–751, Apr. 2016.
[5] C. Gao, J. Zeng, and Z. Zhou, "Identification of multiscale nature and multiple dynamics of the blast furnace system from operating data," *AIChE J.*, vol. 57, no. 12, pp. 3448–3458, Mar. 2011.
[6] P. Zhou, Y. Lv, H. Wang, and T. Chai, "Data-driven robust RVFLNs modeling of a blast furnace iron-making process using cauchy distribution weighted M-estimation," *IEEE Trans. Ind. Electron.*, vol. 64, no. 9, pp. 7141–7151, Sep. 2017.
[7] Y. Pan, C. Yang, R. An, and Y. Sun, "Robust principal component pursuit for fault detection in a blast furnace process," *Ind. Eng. Chem. Res.*, vol. 57, no. 1, pp. 283–291, Jan. 2018.
[8] C. Gao, J. Chen, J. Zeng, X. Liu, and Y. Sun, "A chaos-based iterated multistep predictor for blast furnace ironmaking process," *AIChE J.*, vol. 55, no. 4, pp. 947–962, Apr. 2009.
[9] R. A. C. Yang, Z. Zhou, L. Wang, and Y. Pan, "Comparison of different optimization methods with support vecto machine for blast furnace multi-fault classification," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 1204–1209, Sep. 2015.

[10] B. Zhou, H. Ye, H. Zhang, and M. Li, "Process monitoring of iron-making process in a blast furnace with PCA-based methods," *Control Eng. Pract.*, vol. 47, pp. 1–14, Feb. 2016.

[11] Y. Chu and C. Gao, "Data-based multiscale modeling for blast furnace system," *AIChE J.*, vol. 60, no. 6, pp. 2197–2210, Mar. 2014.

[12] Y. Pan, C. Yang, R. An, and Y. Sun, "Fault detection with improved principal component pursuit method," *Chemometric Intell. Lab. Syst.*, vol. 157, pp. 111–119, Oct. 2016.

[13] E. Jove, J. Casteleiro-Roca, H. Quintián, J. A. Méndez-Pérez, and J. L. Calvo-Rolle, "A fault detection system based on unsupervised techniques for industrial control loops," *Expert Syst.*, vol. 36, no. 4, Mar. 2019.

[14] M. S. Escobar, H. Kaneko, and K. Funatsu, "On generative topographic mapping and graph theory combined approach for unsupervised non-linear data visualization and fault identification," *Comput. Chem. Eng.*, vol. 98, pp. 113–127, Mar. 2017.

[15] R. An, C. Yang, and Y. Pan, "Unsupervised change point detection using a weight graph method for process monitoring," *Ind. Eng. Chem. Res.*, vol. 58, no. 4, pp. 1624–1634, Jan. 2019.

[16] J. Yu, "A support vector clustering-based probabilistic method for unsupervised fault detection and classification of complex chemical processes using unlabeled data," *AIChE J.*, vol. 59, no. 2, pp. 407–419, Feb. 2013.

[17] P. Spyridon and Y. S. Boutalis, "Generative adversarial networks for unsupervised fault detection," in *Proc. Eur. Control Conf. (ECC)*, Jun. 2018, pp. 691–696.

[18] C. G. Bezerra, B. S. J. Costa, L. A. Guedes, and P. P. Angelov, "An evolving approach to unsupervised and real-time fault detection in industrial processes," *Expert Syst. Appl.*, vol. 63, pp. 134–144, Nov. 2016.

[19] H. Chen and N. Zhang, "Graph-based change-point detection," *Ann. Statist.*, vol. 43, no. 1, pp. 139–176, Feb. 2015.

[20] H. Chen, "Sequential change-point detection based on nearest neighbors," *Ann. Statist.*, vol. 47, no. 3, pp. 1381–1407, Jun. 2019.

[21] E. Musulin, "Spectral graph analysis for process monitoring," *Ind. Eng. Chem. Res.*, vol. 53, no. 25, pp. 10404–10416, Jun. 2014.

[22] V. Torra and Y. Narukawa, "On a comparison between mahalanobis distance and choquet integral: The Choquet–Mahalanobis operator," *Inf. Sci.*, vol. 190, no. 1, pp. 56–63, May 2012.

[23] K. Q. Ye, "Indicator function and its application in two-level factorial designs," *Ann. Statist.*, vol. 31, no. 3, pp. 984–994, Jun. 2003.

[24] J.-S. Zeng and C.-H. Gao, "Improvement of identification of blast furnace ironmaking process by outlier detection and missing value imputation," *J. Process Control*, vol. 19, no. 9, pp. 1519–1528, Oct. 2009.

[25] Z. Ge and Z. Song, "Performance-driven ensemble learning ICA model for improved non-Gaussian process monitoring," *Chemometric Intell. Lab. Syst.*, vol. 123, no. 15, pp. 1–8, Apr. 2013.

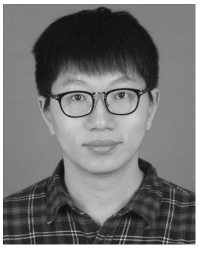
[26] H. Liao, Z. Xu, X.-J. Zeng, and J. M. Merigó, "Qualitative decision making with correlation coefficients of hesitant fuzzy linguistic term sets," *Knowl.-Based Syst.*, vol. 76, pp. 127–138, Mar. 2015.

[27] Z. Zhou, C. Wen, and C. Yang, "Fault isolation based on k-Nearest neighbor rule for industrial processes," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2578–2586, Apr. 2016.

[28] J. Shang, M. Chen, and H. Zhang, "Fault detection based on augmented kernel Mahalanobis distance for nonlinear dynamic processes," *Comput. Chem. Eng.*, vol. 109, pp. 311–321, Jan. 2018.

[29] J. H. Friedman and L. C. Rafsky, "Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests," *Ann. Statist.*, vol. 7, no. 4, pp. 697–717, Jul. 1979.

[30] C. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring," *IFAC Proc. Volumes*, vol. 41, no. 2, pp. 7889–7894, Jul. 2009.



RUQIAO AN received the B.E. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2013, where he is currently pursuing the Ph.D. degree in control science and engineering. His current research interest is data-driven fault diagnosis and its applications in industry.

CHUNJIE YANG (Senior Member, IEEE) is currently a Professor with the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China. His research interests include system modeling, control and fault diagnosis of industrial processes, soft sensor technology, and implementation for complex industrial systems.

YIJUN PAN received the B.E. degree in automation from Jilin University, Changchun, China, in 2013, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2018. She is currently an Assistant Researcher with the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China. Her current research interests are data-driven fault diagnosis and its applications in industry.