Received September 17, 2017, accepted October 17, 2017, date of publication October 25, 2017, date of current version December 5, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2766235

Joint Data-Driven Fault Diagnosis Integrating Causality Graph With Statistical Process Monitoring for Complex Industrial Processes

JIE DONG, MENGYUAN WANG, XIONG ZHANG, LIANG MA, AND KAIXIANG PENG, (Member, IEEE)

Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

Corresponding author: Kaixiang Peng (kaixiang@ustb.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61473033 and Grant 61773053, in part by the Fundamental Research Funds for the China Central Universities of USTB, China, under Grant FRF-BD-16-005A, and in part by the Newton Research Collaboration Award of the Royal Academy of Engineering, U.K., under Grant R33265.

ABSTRACT In this paper, an integrated fault diagnosis method is proposed to deal with fault location and propagation path identification. A causality graph is first constructed for the system according to the *a priori* knowledge. Afterward, a correlation index (CI) based on the partial correlation coefficient is proposed to analyze the correlation of variables in causality graph quantitatively. To achieve accurate fault detection results, the proposed CI is monitored by probability principal component analysis. Moreover, the concept of weighted average value is introduced to identify fault propagation path based on reconstruction-based contribution and causality graph after detecting a fault. Finally, the new proposed scheme would be practiced with real industrial HSMP data, where the individual steps as well as the complete framework were extensively tested.

INDEX TERMS Joint data-driven, fault location, propagation path identification, causality graph, PPCA.

I. INTRODUCTION

Fault diagnosis is essential to modern industrial processes to keep the process within a controllable safety operating region. Thanks to the development of new networked instruments and sensor technology, a large amount of process data can be collected and stored. Consequently, the data-driven based fault diagnosis methods become mainstream technology in process monitoring [1]–[5]. Multivariate statistical process monitoring (MSPM) techniques have been developed to extract useful information from a large number of highly correlated process variables. Probabilistic principal component analysis (PPCA) [6] is one of the most widely used methods in the MSPM field. It can handle incomplete data well, even data with some missing values by using the conditional probability density of the variables [7]. Simultaneously, the probabilistic model can be determined by the expectation maximization (EM) algorithm, which is easy to combine with other methods as a mixture model [8]–[10]. However, lacking analysis of the relationships between process variables, the propagation paths of faults cannot be identified.

For online fault diagnosis, contribution plot and its improved methods have become popular diagnostic tools to identify faulty variables [11]. The contribution plot has been firstly presented to isolate fault variables by comparing the contribution of each process variable to the monitoring statistics [12], [13]. However, contribution plots suffer from the smearing effect in many situations, which leads to erroneous results. Therefore, some further discussion has been made on the application and development of contribution plots [14], [15]. The so-called reconstruction-based contribution (RBC) method has been put forward [16], which combines contribution analysis and reconstruction based identification together. In addition, the method has been extended to general process faults, which is proposed to diagnose the fault type for output-relevant faults [17], [18]. Furthermore, the weighted RBC has been presented to reduce fault smearing and improve diagnosis accuracy [19].

To obtain a complete fault propagation path, causality analysis is introduced. Causality is the relation between the cause and the effect [20], [21]. Many researches on causality analysis have been studied [22]. Among the most commonly used methods are the Granger causality (GC) [23], [24], the transfer entropy [25], [26], cross-correlation analysis [27] and the dynamic Bayesian networks (DBN) [28], [29], which are all based on historical process data and does not require prior information on the intrinsic system. Typically, the outcome of causal analysis is a causal model representing process variables as nodes and causal relationships as arrows. Causality graph is an uncertainty reasoning method proposed by Zhang [30]. The causality of process variables is expressed clearly and intuitively in the graph [31]. Causality graph can be developed using mathematical equations describing the system or directly from piping and instrumentation diagrams (P&IDs). The method has rapid development in the late years and has many expansions, such as multi value causality graph [32], continuous causality graph [33], etc. Owing to its advantages and development, the casuality graph method has been highly applied in fault diagnosis field [34], [35].

Generally, three kinds of methods, namely, graph theory, expert system and qualitative simulation are used for fault diagnosis due to the combination of process knowledge, especially the fault knowledge and related deductions. The graph theory with a simple modeling procedure can get comprehensible diagnosis results with the utilization of sign directed graph (SDG) [29], causality graph [36], [37], bond graph [38] and fault tree analysis [39], etc. However, for complex systems, the search process is complicated with low accuracy, which leads to invalid fault diagnosis results. In a way, the knowledge based fault diagnosis methods can be regarded as the extension of data-driven methods, thus belonging to the field of data-driven fault diagnosis [40], mainly because large amount of historical and real-time data is needed for real-time monitoring and fault diagnosis. Additionally, the method can realize fault diagnosis through various information such as expert or process knowledge, abnormal conditions, fault features and system operation constraints based on causality analysis, fault tree analysis, rules or case-based reasoning, etc. Consequently, on the basis of data and knowledge, more complete fault diagnosis results from quantitative and qualitative analysis can be acquired with the joint data-driven fault diagnosis methods.

In this paper, the knowledge based fault diagnosis method is combined with data-driven method to identify a complete fault propagation path. More concretely, based on causality graph and PPCA method, an integrated fault diagnosis method using RBC is proposed. The main contribution of this paper is summarized as follows. 1) The proposed correlation index (CI) can quantitatively measure correlation of variables. With the combination with causality graph, the causality, connectivity and correlation are all captured. 2) Based on RBC, the weighted average value of vector is introduced to obtain the most possible fault propagation path among all possible paths.

The remainder of this paper is organized as follows. In Section II, the calculation procedure of CI and related definitions are provided. The process monitoring of CI based on PPCA fault detection method is given in Section III. The proposed integrated fault diagnosis method is stated in Section IV. To demonstrate the validity of our approach, fault diagnosis of industrial HSMP are considered in Section V. Finally, the conclusions are given in Section VI.

II. CORRELATION ANALYSIS OF PROCESS VARIABLES

A. CORRELATION ANALYSIS AND RELEVANT DEFINITIONS

Correlation analysis is a statistical method expressing the relationship of process variables. It means that a variable will change as others change. The statistics to measure correlation is called the correlation coefficient. The degree of correlation is represented by the number ranging from 0 to 1. '0' represents un-correlation while '1' stands for full-correlation. The larger value means the stronger correlation.

In correlation analysis, one of the most impressive study is Pearson correlation coefficient [41]. It has been proven to be efficient and widely used into practice. Nevertheless, there are still some drawbacks. Deviations from normal conditions may be quite large if the variables do not follow Gaussian distribution. In addition, the calculation results can be easily affected by abnormal points and the effect is usually significant.

Generally, Pearson correlation coefficient describes the correlation between two variables, while the variables in systems are always correlative with more than two variables. Therefore, multivariate correlation coefficient should be taken into consideration. It consists of the following three concepts : partial correlation coefficient between two variables regardless of the influence of the others, multiple correlation coefficient of a variable to multiple variables and canonical correlation coefficient of multiple variables to multi variables. In this paper, based on causality graph, the correlation between two variables in the graph should be studied. Obviously, partial correlation coefficient is more appropriate, which can express the essential relation between variables more accurately and reliably.

B. THE EXTRACTION OF CORRELATION INDEX

For the purpose of more intuitive and accurate relationships between variables, namely a quantitative description of correlation, the correlation index (CI) is presented based on partial correlation coefficient [42] in this paper. Specifically, for the variables x and y, the correlation coefficient r_{xy} can be calculated as

$$r_{xy} = \frac{cov(x, y)}{\sigma_x \sigma_y} = \frac{E\left[(x - u_x)(y - u_y)\right]}{\sigma_x \sigma_y} \tag{1}$$

where u_x, u_y, σ_x, σ_y are the mean and standard deviation of variables, respectively. To analyze the correlation between two specific variables, Eq.1 can be extended to the matrix

$$r = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \tag{2}$$

Then the inverse matrix c is

$$c = r^{-1} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \tag{3}$$

The partial correlation coefficient between two variables can be expressed as

$$p_{ij} = -\frac{c_{ij}}{\sqrt{c_{ii} \times c_{ij}}} \tag{4}$$

whose corresponding partial correlation coefficient matrix is

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{5}$$

Based on Eq.5, we propose a correlation index CI

$$CI = (v - \bar{v})^T P^{-1} (v - \bar{v}) \tag{6}$$

where $v = \begin{bmatrix} x & y \end{bmatrix}^T$, $\bar{v}$ is the mean value of v, P^{-1} is the inverse matrix of the partial correlation coefficient matrix P.

CI is presented to incorporate multivariate statistics with correlation information. The basic idea of CI is to measure deviation of the correlation from real-time abnormal conditions and historical normal conditions. Besides a quantitative description of correlation, another property of CI is improving the monitoring effect. It is said that better fault monitoring results are achieved for features that are more application-dependent [43]. In terms of application-dependence of the feature generation, features that measure the linkages between process variables are more appropriate than features directly using the process variables themselves for the monitoring of process correlation structures [35].

III. MONITORING OF CI BY USING PPCA FOR FAULT DETECTION

Probabilistic principal component analysis (PPCA) is one of the topical issues in process monitoring. It treats those non principal components as Gaussian noise to further estimate. In addition, both principal components and noise variables in PPCA are monitored via Mahalanobis norm, which overcomes the flaw that two different statistics may cause inconsistent monitoring results in PCA. In this paper, PPCA is applied to monitor CI for fault detection.

A. PRE-PROCESSING OF DATA

In PPCA, data is required to follow or approximately follow Gaussian distribution. However, CI in Eq.6 does not follow Gaussian distribution and its solutions are positive. Thus, it is necessary to preprocess data with data transformation conversion methods, such as Box-Cox transformation [44]. It is actually a family of power transformations that incorporates and extends the traditional options to find the optimal normalizing transformation for each variable. The transformation formula is as follows

$$X^{(\lambda)} = \begin{cases} \frac{X^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \ln X & \lambda = 0. \end{cases} \tag{7}$$

The determination of λ is the key to data transformation with two commonly used methods, maximum likelihood estimation and Bayes method. Also, the Box-Cox transformation function in Minitab software can obtain a right λ. The paper briefs maximum likelihood estimation method here. The formula is written as

$$\begin{aligned} L_{\max} &= -\frac{1}{2} \ln \hat{e}^2 + \ln J(\lambda, X) \\ &= -\frac{1}{2} \ln \hat{e}^2 + (\lambda - 1) \sum_{i=1}^{n} \ln X_i \end{aligned} \tag{8}$$

where for all λ, $J(\lambda, X) = \prod_{i=1}^{n} \frac{\partial W_i}{\partial X_i} = \prod_{i=1}^{n} X_i^{\lambda-1}$. For each λ, $\hat{e}^2$ which can be obtained from $\hat{e}^2 = \frac{1}{2}\sum_{i=1}^{n}\left(X_i^{(\lambda)} - \bar{X}^{(\lambda)}\right)^2$ is the maximum likelihood estimation of $X^{(\lambda)}$. On this basis, $L_{\max}^{(\lambda)}$ can be calculated and λ^* corresponding to the largest $L_{\max}^{(\lambda^*)}$ is the solution.

B. CONSTRUCTION OF PPCA MODEL

For observation vector $x \in R^m$, it can be expressed as the following equation:

$$x = Wt + u + \varepsilon \tag{9}$$

where $t \in R^A$ is the principal component variable. $W \in R^{m \times A}$ $(A < m)$ is the load matrix. u is the mean value of x and ε is the noise following Gaussian distribution. In general, $t \sim N(0, I)$, $\varepsilon \sim N\left(0, \sigma^2 I\right)$, where σ^2 is the variance of the noise. Based on the characters of Gaussian distribution, there exists

$$x|t \sim N\left(Wt + u, \sigma^2 I\right) \tag{10}$$

According to Bayes' theorem, the observation vector x follows

$$x \sim N(u, C) \tag{11}$$

where $C = WW^T + \sigma^2 I$, $C \in R^{m\times m}$, W, σ^2 are parameters to be determined. The probability distribution function of principal component variable t can be written as

$$p(t) = (2\pi)^{-A/2} e^{-\frac{1}{2}t^T t} \tag{12}$$

The conditional probability distribution function of x under condition t is

$$p(x|t) = \left(2\pi\sigma^2\right)^{-m/2} e^{-\frac{1}{2\sigma^2}\|x-Wt-u\|^2} \tag{13}$$

Accordingly, the probability distribution function of x is

$$\begin{aligned} p(x) &= \int p(x|t)\, p(t)dt = (2\pi)^{-m/2}|C|^{-1/2}e^{\zeta} \\ \zeta &= -\frac{1}{2}(t-u)^T C^{-1}(t-u). \end{aligned} \tag{14}$$

On the basis of Bayes probability formula, the posterior distribution of t on x is

$$\begin{aligned} p(t|x) &= (2\pi)^{-\frac{A}{2}} \left|\sigma^{-2}M\right|^{1/2} e^{-\frac{1}{2}\eta(\sigma^{-2}M)\xi} \\ \eta &= t - M^{-1}W^T(x-u)^T \\ \xi &= t - M^{-1}W^T(x-u) \end{aligned} \tag{15}$$

where $M = W^T W + \sigma^2 I$, $M \in R^{A\times A}$. So far, the PPCA model has been established.

Parameters in PPCA model can be solved by expectation maximization (EM). Latent variable t_i is taken as 'incomplete' data. The complete data consists of observation vector and latent variables [8]. Firstly, the posterior distribution of latent variable t_i on observation variable x_i, $p(t_i|x_i)$ is calculated in E-step. In M-step, W and σ^2 are obtained by maximizing the expectation of its log function of 'complete' data. Finally, the parameters are estimated to be

$$\tilde{W} = SW\left(\sigma^2 I + M^{-1}W^T SW\right)^{-1} \tag{16}$$

$$\tilde{\sigma}^2 = \frac{1}{m} tr\left(S - SWM^{-1}\tilde{W}^T\right) \tag{17}$$

where S is the covariance matrix of samples. The equations above should be iterated till convergence and the parameters are obtained.

C. CONSTRUCTION OF PPCA MODEL

For a new sample x_{new}, its latent variable can be calculated by

$$t_{new} = Qx_{new} = W^T\left(WW^T + \sigma^2 I\right)^{-1} x_{new} \tag{18}$$

where $Q = W^T\left(WW^T + \sigma^2 I\right)^{-1}$. The variance of latent variable is $\mathrm{var}(t_{new}) = Q\left(WW^T + \sigma^2 I\right)Q^T$, which indicates that it has no connection with the current sample. So the T^2 statistics in principle subspace can be extended as

$$GT^2_{new} = t^T_{new}\,\mathrm{var}(t_{new})^{-1} t_{new} \le GT^2_{\lim}. \tag{19}$$

Similar to PCA, the SPE statistics is solved as $GSPE_{new} = e^T_{new}\left(\sigma^2 I\right)^{-1} e_{new} \le GSPE_{\lim}$ in noise subspace, where $e_{new} = x_{new} - Wt_{new} = (I - WQ)\,x_{new}$ is the error of the new sample. The two statistics follow χ^2 distribution of appropriate freedom:

$$\begin{aligned} GT^2_{\lim} &= \chi^2_{1-\alpha}(A) \\ GSPE_{\lim} &= \chi^2_{1-\alpha}(m) \end{aligned} \tag{20}$$

where α is the significance and $1-\alpha$ represents the credibility. A and m are the freedom, respectively, which can be adjusted in practice for a better detection results.

In PPCA, the statistics in principle subspace and noise subspace can be integrated by using the Mahalanobis norm. A comprehensive monitoring index ST is given by Ghahramani *et al.* [8], which can be used independently in process monitoring and defined as

$$ST = \left\|C^{-\frac{1}{2}} x_{new}\right\|^2 = x^T_{new} C^{-1} x_{new} \le ST_{\lim} \tag{21}$$

where $ST_{\lim} = \chi^2_{1-\alpha}(m)$.

D. MONITORING OF CI VIA PPCA

As mentioned earlier, CI is a quantitative description of the correlation which measures the linkages between process variables. Once the linkages are broken, namely, CI exceeding the predefined threshold, there may be some faults in the system under consideration. To achieve successful fault detection of industrial processes, the proposed CI is monitored with the utilization of PPCA based fault detection method. According to the two previous sections, CI index can be written as

$$x = Wt + u + \varepsilon. \tag{22}$$

Similarly, the parameters of PPCA based on the training CIs are able to be solved using EM algorithm.

For a new CI obtained under real-time condition CI_{new}, its latent variable is

$$t_{CInew} = QCI_{new} = W^T\left(WW^T + \sigma^2 I\right)^{-1} CI_{new}. \tag{23}$$

Thus, the T^2 statistics GT^2_{CInew} and SPE statistics $GSPE_{CInew}$ can be calculated. The so called ST statistics is expressed as

$$ST = \left\|C^{-\frac{1}{2}} CI_{new}\right\|^2 = CI^T_{new} C^{-1} CI_{new}. \tag{24}$$

The correlation of the process is considered normal if the monitoring statistics are below the thresholds, i.e., $GT^2_{CInew} \le GT^2_{CI\lim}$ and $GSPE_{CInew} \le GSPE_{CI\lim}$, $ST_{CI} \le ST_{CI\lim}$. Otherwise, a fault is thought to occur.

IV. INTEGRATED FAULT DIAGNOSIS BASED ON CASUALITY GRAPH AND STATISTICAL PROCESS MONITORING

A. PRELIMINARIES OF CAUSALITY GRAPH

Causality graph theory is a methodology to deal with knowledge representation and reasoning of uncertain causalities based on belief networks proposed by Zhang [30]. At present, it has been developed into a hybrid causality diagram which is capable for discrete variables and continuous variables. Causality graph has the following advantages. 1) It is completely based on probability theory with good theoretical basis. 2) It is able to deal with loop structure. 3) The direct causal intensity rather than probability makes it easy to acquire expert knowledge and experience. 4) Due to its dynamic characteristics, the causal structure is able to change as the information accepted online. 5) Is has flexible modes of reasoning: from causes to consequence, $\Pr\{X|Causes\}$; from consequence to causes, $\Pr\{X|Consequence\}$; hybrid causes and consequence, $\Pr\{X|Causes\ \&\ Consequence\}$.

The directional and structural relationships between variables are described clearly in the graph with nodes representing events or variables. A directed edge stands for a causal relationship with its strength of connection indicating the strength of causality. This graphical representation of knowledge is very intuitive and natural, which is easy to express explicit knowledge and give expert knowledge.

From the characteristics of causality graph, the knowledge expression of causality graph has a good correspondence with fault features of the system. It can effectively express the fault knowledge of complex system and has valid reasoning algorithm. Consequently, the utilization of causality graph in fault diagnosis is of great significance to reduce the time of fault judgement, improve fault detection accuracy and save maintenance costs. In this paper, the qualitative causality graph is firstly acquired from the priori knowledge. The quantitative analysis will be made based on CI, which is the foundation of fault propagation path identification.

B. THE RBC BASED FAULT DIAGNOSIS

The basic idea of contribution plot is to identify faults by the contribution of each variable to SPE and T^2 statistics. Variables with large contribution are the possible fault variables. Nevertheless, the cause of faults needs to be further analyzed and identified by the relevant process knowledge.

In PPCA, the contribution of variables to ST statistics is defined as

$$Cont_i^{ST} = \left(\xi_i^T C^{-1/2} x\right)^2 \tag{25}$$

where $C = WW^T + \sigma^2 I$, ξ_i is the direction vector of variables. When a fault occurs, x can be written as $x = x^* + f_i\xi_i$, where x^* is the fault free part and $f_i\xi_i$ is the fault part with f_i denoting the fault amplitude.

Reconstruct the fault variable as $z_i = x - f_i\xi_i$, so the monitoring index of z_i is $Index\,(z_i) = z_i^T C^{-1} z_i$. By minimizing the index with partial least squares (PLS), f_i is solved as $f_i = \left(\xi_i^T C^{-1}\xi_i\right)^{-1}\xi_i^T C^{-1}x$.

In RBC method, variables with large reconstruction contribution are likely to have faults. The equation of reconstruction contribution is

$$\begin{aligned} Cont_i^{RBC} &= x^T C^{-1}\xi_i\left(\xi_i^T C^{-1}\xi_i\right)^{-1}\xi_i^T C^{-1}x \\ &= \frac{\left(\xi_i^T C^{-1}x\right)^2}{\xi_i^T C^{-1}\xi_i} = \frac{\left(\xi_i^T C^{-1}x\right)^2}{c_{ii}} \end{aligned} \tag{26}$$

where c_{ii} is the elements of ith principal diagonal in matrix C^{-1}.

C. FAULT PROPAGATION PATH IDENTIFICATION BASED ON WEIGHTED AVERAGE VALUE

The fault diagnosis method based on causality graph and RBC may get multiple fault propagation paths, which is inconvenient for fault identification, fault isolation and maintenance. To this end, weighted average value of vector is adopted in this paper. The weighted average value of reconstruction contribution in each path is calculated to determine the most possible fault propagation path and achieve the target of fault location.

For a set of data consisting of a single variable, the weighted average value is

$$L = \sum_{i=1}^{n} k_i C_i \Big/ \sum_{i=1}^{n} k_i \tag{27}$$

where k_i is the weighted coefficient. Similarly, for a set of data consisting of vectors, the weighted average value of vector [45] is defined as: s set of m-dimensional observation vectors is given as $L_1,\ L_2,\ \cdots,\ L_i,\ \cdots,\ L_n$ and the corresponding weighted matrix is $K_1,\ K_2,\ \cdots,\ K_i,\ \cdots,\ K_n$, $(i = 1, 2, \cdots, n)$. So the weighted average value of the observation vectors is

$$\bar{L} = (K_1 + K_2 \cdots + K_n)^{-1}\,(K_1L_1 + K_2L_2 \cdots + K_nL_n)\,. \tag{28}$$

Obviously, $\bar{L}$ is also a vector, which means the weighted average value of vector is the extension of general weighted average value. In order to identify the most possible fault propagation path, the weighted average value of each fault propagation path needs to be calculated. Furthermore, its 2-norm can be acquired. The larger the 2-norm, the greater the possibility of fault propagation.

D. FAULT DIAGNOSIS INTEGRATING CAUSALITY GRAPH WITH STATISTICAL PROCESS MONITORING

In this article, the causality between process variables are first analyzed based on causality graph. In addition, the innovative correlation index CI is proposed to measure the correlation quantitatively, thus capturing causality and correlation simultaneously. PPCA is then adopted to monitor CI and detect faults. After a fault has been detected, the RBC is combined with causality graph and CI to get the possible fault propagation paths. Finally, the weighted average value of vector is introduced to find a most possible one, which attains the goal of fault location and identification.

According to the discussion above, the whole scheme is summarized in Fig.1.

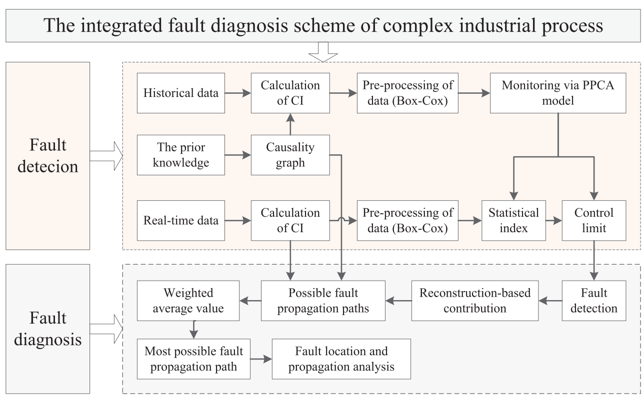

FIGURE 1. Fault diagnosis scheme based on causality graph and statistical process monitoring.

V. APPLICATION IN INDUSTRIAL HOT STRIP MILL

Hot strip mill process is a complex industrial production process, mainly including reheating furnaces, rough rolling, finishing rolling, laminar cooling and coiling etc., as shown in Fig.2. The reheating furnace ensures that the rolling temperature can reach 1200 degrees Celsius when beginning rough rolling. Generally, the strip steel can be rolled into the transfer bar of 28-45mm in the rough rolling zone. Then, it will be sent to the finishing rolling which consists of 6 or 7 stands. The finishing rolling gives further and more precise

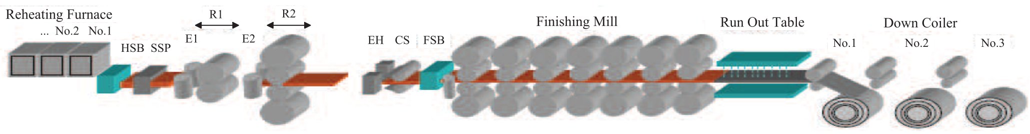

FIGURE 2. The equipment layout of HSMP.

TABLE 1. Variables of hot strip rolling process.

Variables	Type	Description	Unit
G_1-G_7	Process manipulated variables	Average roll gap of i^{th} stand ($i = 1, \ldots, 7$)	mm
F_1-F_7	Process manipulated variables	Rolling force of i^{th} stand ($i = 1, \ldots, 7$)	MN
B_2-B_7	Process manipulated variables	Bending force of i^{th} stand ($i = 2, \ldots, 7$)	MN
S	Quality variables	Shape of finishing mill	mm

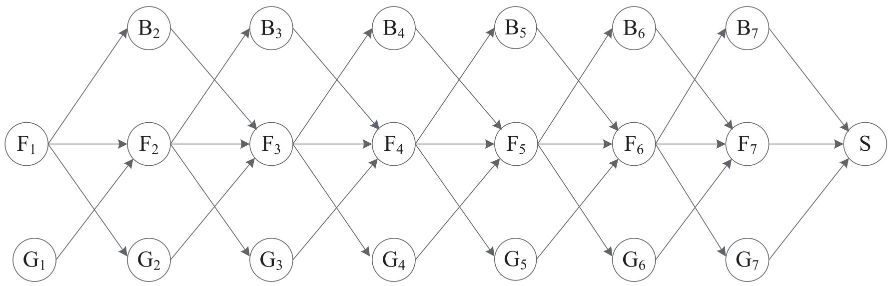

FIGURE 3. Causality graph of variables in HSMP.

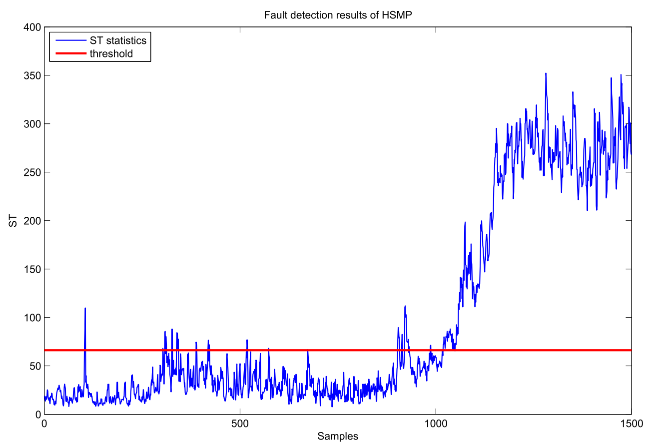

FIGURE 4. Fault detection results of HSMP.

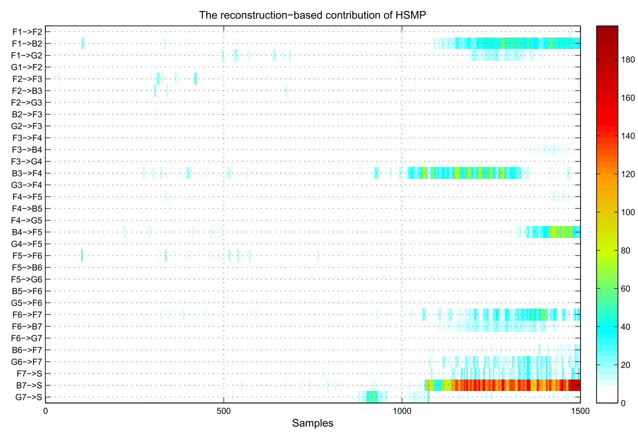

FIGURE 5. RBC of variables in HSMP.

gauge reduction to achieve an accurate needed thickness, which will serve as a background process in this section. Each stand is mainly composed of a machine frame arch, a pair of working rolls, a pair of supporting rolls and the corresponding hydraulic screw-down devices and bending rolls, etc.. The rolling force testing device is installed under the lower supporting rolls. Roll gap between upper and lower working rollers is mainly controlled by the hydraulic servo system in order to meet the requirements of exit thickness of the strip steels. The thickness gauge, thermometric indicator, width gauge and shape meter etc. are arranged at the finishing mill exit. The rolling mill control system is generally equipped with automatic gauge control (AGC), finishing temperature control (FTC), automatic shape control (ASC) and so on, in order to achieve the corresponding requirements of strip thickness, temperature and shape. Eventually, after laminar cooling, the strip steels are coiled as final products.

In HSMP, thickness, width, shape and exit temperature can affect the quality of products. In this paper, 1700mm strip hot rolling production line of Ansteel Corporation in

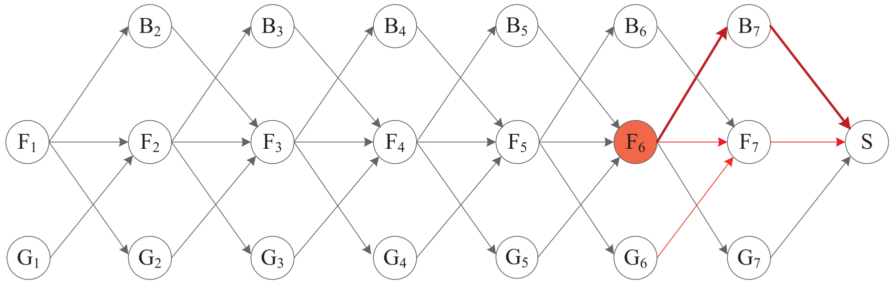

FIGURE 6. Fault location and propagation path identification in HSMP.

Liaoning Province China is applied to the achieved results. 20 process variables including average roll gap, rolling force and bending force (except the first stand) of 7 stands are taken into consideration. The shape of finishing mill is regarded as quality variable. The specific variable information in the process is listed in Table 1. As the shape is required to fluctuate within a range. Thus, the output is similar to follow Guassian distribution. Consequently, based on the historical data set, the PPCA model could be established with 20 measured process variables and one quality variable. The causality graph of variables in HSMP is given in Fig.3 based on the process mechanism. In this case, an actuator failure of cooling water control valve occurs between the fifth and sixth stand. Thus the rolling force of the bracket behind the sixth stand is changed. With the feedback regulation of AGC, the bending force of the seventh stand is influenced, which results in the change of strip shape. For case study, the sampling interval is set to be 10ms. 2000 normal historic data are used for modeling, and 1500 real-time sampling data are used for the testing step. Accordingly, the fault detection results are shown in Fig.4. The statistics exceed control limit evidently at 550th sample, which indicates a fault has occurred.

The RBC of variables in HSMP is provided in Fig.5. According to the contribution and causality graph, the fault location and possible fault propagation paths can be determined. With the utilization of weighted average value of vectors, the most possible propagation path is acquired and shown in Fig.6. It can be seen that the rolling force of 6th stand has been changed, which influences the bending force of 7th stand and further affects the shape of strip steel. The diagnostic result is consistent with the practical situation, which verifies the validity of the proposed method. The timely maintenance can be implemented, thus reducing the costs.

VI. CONCLUSION

In the paper, an integrated fault diagnosis method based on causality graph and statistical process monitoring is proposed. First of all, a causality graph is constructed on the basis of the priori knowledge. Subsequently, a correlation index CI is put forward to quantitatively analyze the relationship of variables in the causality graph. With this combination, the causality, connectivity and correlation between variables are all captured. To improve the effect of fault detection, the Box-Cox transformation is implemented on CI. After a detecting a fault, the possible fault propagation paths are identified with the utilization of RBC and causality graph. Finally, the weighted average value of vector is introduced to get a most possible fault propagation path. The effectiveness of our proposed method is verified by HSMP. Compared with the practical situations, the achieved results is helpful for operators to implement maintenance in time. For future research works, the fault propagation of quality related variables in complex industrial processes will be studied.

ACKNOWLEDGMENT

The authors would like to acknowledge the cooperated Project C108-Z-08-001 with Ansteel Corporation in Liaoning Province, PR China, for providing the data.

REFERENCES

[1] 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.

[2] K. X. Peng, L. Ma, and K. Zhang, "Review of quality-related fault detection and diagnosis techniques for complex industrial processes," *Acta Auto. Sincia*, vol. 43, no. 3, pp. 349–365, 2017.

[3] S. Yin, G. Wang, and H. Gao, "Data-driven process monitoring based on modified orthogonal projections to latent structures," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 4, pp. 1480–1487, Jul. 2016.

[4] B. Shen, Z. Wang, and X. Liu, "Bounded H_∞ synchronization and state estimation for discrete time-varying stochastic complex networks over a finite horizon," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 145–157, Jan. 2011.

[5] H. Zhang, X. Tian, and X. Deng, "Batch process monitoring based on multiway global preserving kernel slow feature analysis," *IEEE Access*, vol. 5, pp. 2696–2710, 2017.

[6] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 61, no. 3, pp. 611–622, 1999.

[7] W. Shi *et al.*, "Temporal dynamic matrix factorization for missing data prediction in large scale coevolving time series," *IEEE Access*, vol. 4, pp. 6719–6732, 2016.

[8] Z. Ghahramani and M. I. Jordan, "Supervised learning from incomplete data via an EM approach," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 6. 1995, pp. 120–127.

[9] J. L. Zhu, Z. Q. Ge, and Z. H. Song, "Robust modeling of mixture probabilistic principal component analysis and process monitoring application," *AIChE J.*, vol. 60, no. 6, pp. 2143–2157, 2014.

[10] T. A. Nakamura *et al.*, "Adaptive fault detection and diagnosis using parsimonious Gaussian mixture models trained with distributed computing techniques," *J. Franklin Inst.*, vol. 354, no. 6, pp. 2543–2572, 2017.

[11] J. Wang, W. Ge, J. Zhou, H. Wu, and Q. Jin, ‘‘Fault isolation based on residual evaluation and contribution analysis,’’ *J. Franklin Inst.*, vol. 354, no. 6, pp. 2591–2612, 2017.

[12] J. F. MacGregor and T. Kourti, ‘‘Statistical process control of multivariate processes,’’ *Control Eng. Pract.*, vol. 3, no. 3, pp. 403–414, 1995.

[13] P. Miller, R. E. Swanson, and C. E. Heckler, ‘‘Contribution plots: A missing link in multivariate quality control,’’ *Appl. Math. Comput. Sci.*, vol. 8, no. 4, pp. 775–792, 1998.

[14] J. A. Westerhuis, S. P. Gurden, and A. K. Smilde, ‘‘Generalized contribution plots in multivariate statistical process monitoring,’’ *Chemom. Intell. Lab. Syst.*, vol. 51, no. 1, pp. 95–114, 2000.

[15] A. Alawi *et al.*, ‘‘Sensor fault identification using weighted combined contribution plots,’’ *IFAC Proc. Volumes*, vol. 39, no. 13, pp. 908–913, 2006.

[16] C. R. Alvarez, A. Brandolin, and M. C. Sánchez, ‘‘On the variable contributions to the *D*-statistic,’’ *Chemom. Intell. Lab. Syst.*, vol. 88, no. 2, pp. 189–196, 2007.

[17] C. F. Alcala and S. J. Qin, ‘‘Reconstruction-based contribution for process monitoring,’’ *Automatica*, vol. 45, no. 7, pp. 1593–1600, Jul. 2009.

[18] L. Gang, C. F. Alcala, S. J. Qin, and Z. Donghua, ‘‘Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the Tennessee Eastman process,’’ *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 5, pp. 1114–1127, Sep. 2011.

[19] H. Xu *et al.*, ‘‘Weighted reconstruction-based contribution for improved fault diagnosis,’’ *Ind. Eng. Chem. Res.*, vol. 52, no. 29, pp. 9858–9870, 2013.

[20] D. Karnopp, ‘‘An approach to derivative causality in bond graph models of mechanical systems,’’ *J. Franklin Inst.*, vol. 329, no. 1, pp. 65–75, 1992.

[21] F. Yang *et al.*, *Capturing Connectivity and Causality in Complex Industrial Processes*. Berlin, Germany: Springer, 1992.

[22] 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.

[23] J. Geweke, ‘‘Measurement of linear dependence and feedback between multiple time series,’’ *J. Amer. Stat. Assoc.*, vol. 77, no. 378, pp. 304–313, Jun. 1982.

[24] A. Porta, T. Bassani, V. Bari, G. D. Pinna, R. Maestri, and S. Guzzetti, ‘‘Accounting for respiration is necessary to reliably infer Granger causality from cardiovascular variability series,’’ *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 832–841, Mar. 2012.

[25] H. Gharahbagheri, S. Imtiaz, F. Khan, and S. Ahmed, ‘‘Causality analysis for root cause diagnosis in Fluid Catalytic Cracking unit,’’ *IFAC-Papers OnLine*, vol. 48, no. 21, pp. 838–843, 2015.

[26] M. Bauer and N. F. Thornhill, ‘‘A practical method for identifying the propagation path of plant-wide disturbances,’’ *J. Process Control*, vol. 18, nos. 7–8, pp. 707–719, 2008.

[27] J. Yu and M. M. Rashid, ‘‘A novel dynamic bayesian network-based networked process monitoring approach for fault detection, propagation identification, and root cause diagnosis,’’ *AIChE J.*, vol. 59, no. 7, pp. 2348–2365, 2013.

[28] J. Mori and J. Yu, ‘‘Dynamic Bayesian network based networked process monitoring for fault propagation identification and root cause diagnosis of complex dynamic processes,’’ *IFAC Proc. Volumes*, vol. 46, no. 32, pp. 678–683, 2013.

[29] P. Duan, T. Chen, S. L. Shah, and F. Yang, ‘‘Methods for root cause diagnosis of plant-wide oscillations,’’ *AIChE J.*, vol. 60, no. 6, pp. 2019–2034, 2014.

[30] Q. Zhang, ‘‘Probabilistic reasoning based on dynamic causality trees/diagrams,’’ *Rel. Eng. Syst. Safety*, vol. 46, no. 3, pp. 209–220, 1994.

[31] J. Mori, V. Mahalec, and J. Yu, ‘‘Identification of probabilistic graphical network model for root-cause diagnosis in industrial processes,’’ *Comput. Chem. Eng.*, vol. 71, pp. 171–209, Dec. 2014.

[32] X. Fan, Z. Qin, S. Maosong, and H. Xiyue, ‘‘Reasoning algorithm in multi-value causality diagram,’’ *Chin. J. Comput.*, vol. 26, no. 3, pp. 310–322, 2003.

[33] Q. Zhang, ‘‘Dynamic uncertain causality graph for knowledge representation and reasoning: Discrete DAG cases,’’ *J. Comput. Sci. Technol.*, vol. 27, no. 1, pp. 1–23, 2012.

[34] J. Thambirajah, L. Benabbas, and M. Bauer, ‘‘Cause-and-effect analysis in chemical processes utilizing XML, plant connectivity and quantitative process history,’’ *Comput. Chem. Eng.*, vol. 33, no. 2, pp. 503–512, 2009.

[35] B. Jiang, X. Zhu, D. Huang, and R. D. Braatz , ‘‘Canonical variate analysis-based monitoring of process correlation structure using causal feature representation,’’ *J. Process Control*, vol. 32, pp. 109–116, Aug. 2015.

[36] H. Gharahbagheri, S. A. Imtiaz, and F. Khan, ‘‘Root cause diagnosis of process fault using KPCA and Bayesian network,’’ *Ind. Eng. Chem. Res.*, vol. 56, no. 8, pp. 2054–2070, 2017.

[37] L. H. Chiang, B. Jiang, X. Zhu, D. Huang, and R. D. Braatz, ‘‘Diagnosis of multiple and unknown faults using the causal map and multivariate statistics,’’ *J. Process Control*, vol. 28, pp. 27–39, Apr. 2015.

[38] P. Zhao, W. Wenhui, and Z. Donghua, ‘‘Diagnosis of sensor and actuator faults of a class of hybrid systems based on semi-qualitative method,’’ in *Proc. 5th World Congr. Intell. Control Auto.*, vol. 2. Jun. 2004, pp. 1771–1774.

[39] S. Kabir, ‘‘An overview of fault tree analysis and its application in model based dependability analysis,’’ *Expert Syst. Appl.*, vol. 77, pp. 114–135, Jul. 2017.

[40] Z. W. Gao, C. Cecati, and S. X. Ding, ‘‘A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches,’’ *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015.

[41] K. Person, ‘‘Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia,’’ *Philos. Trans. Roy. Soc. London A, Containing Papers Math. Phys. Character*, vol. 187, pp. 253–318, Jan. 1896.

[42] A. de la Fuente, N. Bing, I. Hoeschele, and P. Mendes, ‘‘Discovery of meaningful associations in genomic data using partial correlation coefficients,’’ *Bioinformatics*, vol. 20, no. 18, pp. 3565–3574, 2004.

[43] M. S. Sarfraz, O. Hellwich, and Z. Riaz, ‘‘Feature extraction and representation for face recognition,’’ in *Face Recognition* (Computer and Information Science). 2010.

[44] J. W. Osborne, ‘‘Improving your data transformations: Applying the Box-Cox transformation,’’ *Pract. Assessment, Res. Eval.*, vol. 15, no. 12, pp. 1–9, 2010.

[45] Y. M. Zhou, ‘‘Weighted average value of vector its application,’’ *Bull. Surv. Mapping*, vol. 2, pp. 11–26, 2004.

JIE DONG received the B.E., M.E., and Ph.D. degrees from the University of Science and Technology Beijing, in 1995, 1997, and 2007, respectively. She is currently an Associate Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing. In 2004, she visited The University of Manchester as a Visiting Scholar. Her research interest covers intelligent control theory and application, process monitoring and fault diagnosis, and complex system modeling and control.

MENGYUAN WANG received the B.E. degree in automation from the University of Science and Technology Beijing, Beijing, China, in 2015, where she is currently pursuing the master's degree in control science and engineering. Her research interests include process monitoring, fault diagnosis, and prediction for industrial processes.

XIONG ZHANG received the B.E. degree in automation from the University of Science and Technology Liaoning, Anshan, China, in 2014. He is currently pursuing the master's degree in control science and engineering with the University of Science and Technology Beijing, Beijing, China. His research interests include process monitoring and fault diagnosis.

LIANG MA received the master's degree in control theory and control engineering from the North China University of Science and Technology, Tangshan, China, in 2012. He is currently pursuing the Ph.D. degree with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China. His research interest covers fault diagnosis and fault tolerant control.

KAIXIANG PENG (M'15) received the B.E. degree in automation and the M.E. and Ph.D. degrees from the Research Institute of Automatic Control, University of Science and Technology Beijing, Beijing, China, in 1995, 2002, and 2007, respectively. He is a Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His research interests are fault diagnosis, prognosis, and the maintenance of complex industrial processes, modeling and control for complex industrial processes, and control system design for the rolling process.