Received May 7, 2020, accepted May 31, 2020, date of publication June 10, 2020, date of current version June 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3001379

A Data-Driven Fault Detection Framework Using Mahalanobis Distance Based Dynamic Time Warping

YULIN SI[1,2], (Member, IEEE), ZHENG CHEN[1], JILI SUN[1], DAHAI ZHANG[1], (Member, IEEE), AND PENG QIAN[3]

[1]Ocean College, Zhejiang University, Zhoushan 316021, China
[2]Key Laboratory of Ocean Observation-Imaging Testbed of Zhejiang Province, Zhoushan 316021, China
[3]Department of Mechanical, Aerospace, and Civil Engineering, Brunel University, London UB8 3PH, U.K.

Corresponding author: Peng Qian (qianpeng18@sina.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51705453, Grant 51879233, and Grant 61911530251, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LHY20E090001, in part by the Zhoushan Municipal Commission of Science and Technology under Grant 2019C81036, and in part by the Fundamental Research Funds for the Central Universities.

ABSTRACT Fault detection module is one of the most important components in modern industrial systems. In this paper, we propose a novel fault detection framework which makes use of both normal and faulty measurement signals at the same time. In this framework, the multivariate time series (MTS) pieces which are extracted from measurement signals in a time interval are used as the training and testing samples, and a K-nearest neighbour rule of MTS pieces is applied for fault detection. Moreover, a Mahalanobis distance based dynamic time warping method is used to measure the divergence among MTS pieces, and a one-class metric learning algorithm is proposed to learn the appropriate Mahalanobis distance. Experimental results on the Tennessee Eastman process demonstrate that the proposed method has improved fault detection performance compared with classical approaches on certain kinds of faults.

INDEX TERMS Data-driven, fault detection, multivariate time series, Mahalanobis distance, dynamic time warping.

I. INTRODUCTION

In modern industrial systems with high degree of automation, process monitoring and fault diagnosis is one of the most important blocks, which plays an important role as it strengthens the safety as well as reliability of the industrial process. Therefore, research on fault detection methods has drawn great attention in recent years in both academia and industry [1]–[4]. Modern industrial systems have become more and more complex with the rapid development of industrial automation degree. Thus it is very difficult to establish an accurate model to detect the occurrence of fault as these system models are always based on human expertise or priori knowledge [5], so it is believed that model-based fault detection approaches [6], [7] can hardly achieve good performance regarding complex industrial systems in practice.

The associate editor coordinating the review of this manuscript and approving it for publication was Tyson Brooks.

In contrast, data-driven fault diagnosis algorithms [8]–[12] do not rely on models from first principles but learning fault detection models directly from the measured data from running systems. Different from model-based approaches, data-driven fault detection methods are used to establish the relationship between the measured signals in fault-free and faulty operations, which do not rely on any human expertise or priori knowledge of the system. As a practical solution, data-driven fault detection approach has drawn great attention from numerous scholars in both research and application fields [13]–[16].

The main objective of data-driven fault detection is to analyse the measurement signals of the system to determine whether any fault has occurred or not [17]. Different from normal operation, a fault is defined as an abnormal behavior, which might be caused by electric failure, component damage, extreme process disturbances etc. In the fault detection process, the measurement signals are collected using various sensors, and the observable and measurable variables, such as temperature, pressure, flow rate, etc. are recorded periodically. These time-varying variables can be used to construct a measurement signal, which is a uniformly sampled multivariate time series (MTS) [18]. When a fault occurs, some or all variables of measurement signals will change, and certain patterns of measurement signals in faulty cases are different from those in normal operation. Finding appropriate measurement signal patterns is the core part of data-driven fault detection methods.

Many data-driven fault detection methods including principal component analysis (PCA) [17], [19], partial least squares (PLS) [20]–[22] and fisher discriminant analysis (FDA) [25] based fault detection methods assume that the normal measurement signals follow multivariate Gaussian distribution. These methods firstly learn the parameters of the multivariate Gaussian distribution model. If the data in the testing measurement signals are outside of the confidence interval, the system are then regarded as faulty. PCA is a statistical method which has high efficiency in processing data with high dimensions. The PCA based fault detection method uses orthogonal transformation to convert the observable data and predicted data into orthogonal components. These orthogonal components are then divided into principal components and less important components, which will build two rules called squared prediction error (SPE) and Hotelling T^2 to determine whether the system is in normal or faulty operation mode. The extended versions, including multi-scale PCA (MSPCA) [19], modified PCA (MPCA) [26] based methods improve its fault detection performance to some degree. Partial least squares(PLS) [20]–[22] is another powerful statistical tool used for fault detection, which utilizes the covariance information of the measurement signals to identify a linear correlation model. The model projects the observable data as well as the predicted data into a new space. PLS based method also uses SPE and Hotelling T^2 as the criteria of fault detection. One modified version of PLS, called modified PLS (MPLS) [22], estimates the correlation model in the least-square sense at first, then uses an orthogonal decomposition process on the measurement space. In the work [21], the authors further decompose the results of standard PLS on certain subspaces to avoid the oblique decomposition [27] on measurement space, which is one weak point of standard PLS based algorithm. This method is named as the total projection to latent structure (TPLS) based fault detection approach. FDA is a dimensionality reduction technique and it is good at discriminating data with different labels. The FDA based fault detection method firstly calculates two matrices (within-class scatter matrix and between class-scatter matrix) using observable data, and then eigenvectors are obtained from these two matrices. The FDA method projects the predicted data on that eigenvectors and determine the label of the data using a T^2 criterion.

Another category of data-driven fault detection methods deal with measurement signals with non-Gaussian distributions. For instance, the independent component analysis (ICA) [28] based fault detection method only requires the measurement signals to be decomposed into a linear combination of non-Gaussian variables, which are called independent components (ICs). ICA based method extracts the hidden statistically ICs from the observed data, and these ICs construct three criteria, which are simultaneously used to monitor the status of the running system. In the work [29], a modified ICA (MICA) based algorithm reduces the computation load of the standard ICA based approach by offering a unique solution of ICs, which improves the fault detection performance. The assumption limitation of ICA based methods are more relaxed than that of the PCA, PLS and FDA based fault detection strategies. That means ICA can be used in wider range of applications. Alternatively, subspace aided approach (SAP) [30] can also deal with measurement signals with non-Gaussian distribution. The main idea of SAP is to identify a primary form of residual generators directly from the observed measurement signals, and a T^2 value is seen as the final criterion in the detection of predicted measurement signals.

Most data-driven fault detection approaches detect faults by processing the static features of samples. These static features are obtained using the statistical value of the measurement signals or special points. However, there are many industrial process control systems that can capture multivariate time varying measurement signals, and only using static features might neglect the dynamic characteristics. One type of dynamic features is the jumping characteristics for certain faults, in which the static features during a range of time intervals are abnormal, but they will return to normal status afterwards while the system is still in faulty operation. This process might only repeat for several time cycles, which will result in a low fault detecting rate. For example, in the famous Tennessee Eastman (TE) process [31], several testing faulty measurement signals given in [32] have jumping characteristics, and it was proven that most classical fault detection methods have low fault detection rate on IDV(11), IDV(16) and IDV(19) due to the influence of the jumping characteristics [33]. To solve this problem, a good idea is to use multivariate time series as the training and testing samples. The authors in [34] proposed a dynamic PCA (DPCA) method which inspects the measurement signals in a time interval, and the following procedures are identical with standard PCA. This method utilizes dynamic features in the fault detection process to some extent. However, DPCA is also based on the assumption that normal measurement signals follow multivariate Gaussian distribution.

Many classical fault detection approaches train the model only by using normal data and determine the threshold boundaries by eliminating the statistical values of normal data [23], [24]. If the testing data are within the scope enveloped by the threshold boundaries, they will be labeled as normal, and vice versa. However, there are many applications that both of the normal and faulty training data can be offered in the off-line design procedure, such as the TE process. In this case, faulty measurement signals are not considered in the

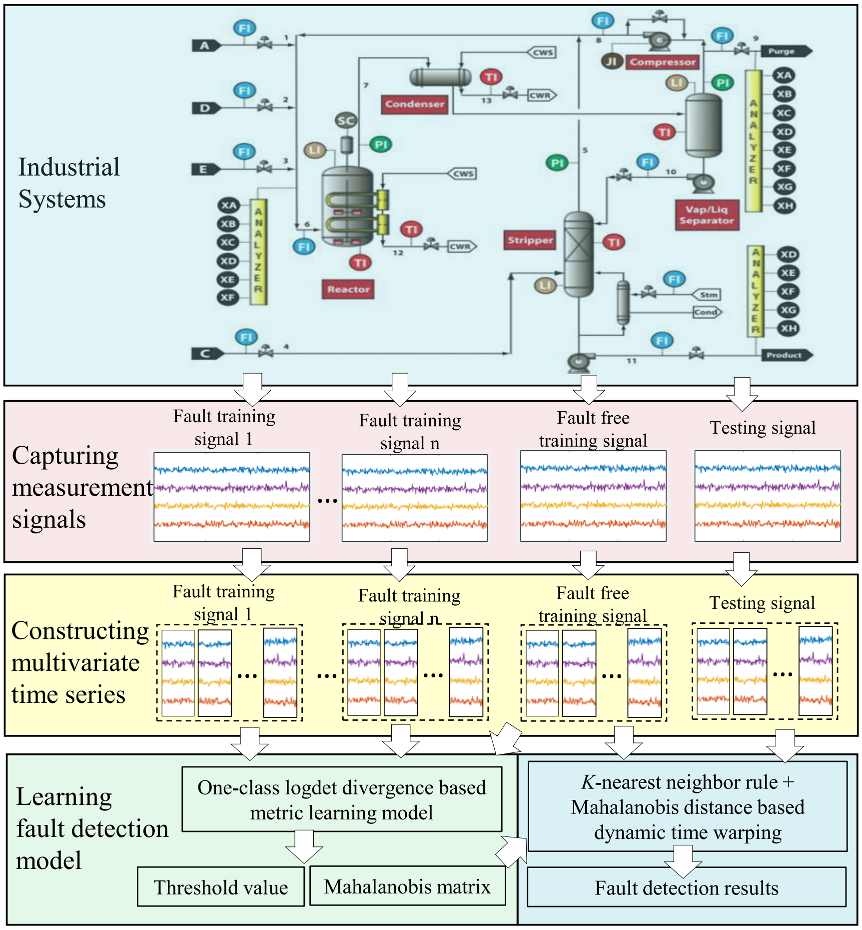


FIGURE 1. **Framework of the proposed fault detection method.**

training process of these fault detection methods, thus this will waste information resources. A good solution is to train a model which can make full use of the normal and faulty signals.

In this paper, we establish a novel data-driven fault detection framework for industrial processes, in which multivariate time series are used to represent the dynamic features of the measurement signals, and a multivariate dynamic time warping method based on Mahalanobis distance is proposed. In order to obtain the Mahalanobis distance function, we propose a one-class metric learning algorithm, which learns a distance metric where the normal samples have concentrated distribution while the faulty samples are far away from normal samples. The distinct boundary between normal and faulty signals helps to improve the fault detection performance. In the end, the TE process is used to verify the proposed data-driven fault detection method, which includes faults that can hardly be detected by traditional methods. The proposed framework is shown in Fig. 1.

The remainder of this paper is organized as follows. In Section II, related literature and background knowledge are presented. Then, the proposed Mahalanobis distance based dynamic time warping algorithm is described in Section III. Section IV illustrates the one-class LogDet divergence based metric learning algorithm for MTS. Section V gives experimental results on TE process to demonstrate the effectiveness of the proposed method. At last, we draw conclusions and point out future directions.

II. RELATED WORKS

In traditional data-driven fault detection methods, SPE and Hotelling T^2 are the most popular indices to determine whether the system is in normal or faulty operation mode. However, these two decision rules both require the measurement signals in normal mode to follow a multidimensional Gaussian distribution. Meanwhile, the faulty measurement signals should be outside of the multi-dimensional Gaussian distribution learnt from training samples in normal mode. However, in many occasions, the normal and faulty measurement signals are hard to be distinguished only by SPE and T^2 learnt from the normal training data. Many industrial processes, such as semiconductor manufacturing processes [35], [36] are nonlinear and multi-modal, where the SPE and the T^2 do not work well. The works [35]–[37] recommended to use nonparametric detection rules which do not require assumption on the probability distribution of normal and faulty measurement signals, e.g. the k-nearest neighbor (KNN) rule.

KNN is a tool widely used in classification problems. An unlabelled sample is classified according to the labels of its nearest neighboring samples. Therefore, the KNN rule can be applied in data with any sort of distribution. Suppose x_i is a training sample with label $y_i \in \{-1, 1\}$, $i = 1, 2, \cdots, n$, the predicted label of a testing sample z using KNN rule can be expressed as

$$h^{\mathrm{KNN}}(z) = \mathrm{sign}\Big(\sum_{i=1}^{n} \eta^k(z, x_i) y_i\Big), \tag{1}$$

where $\eta(\cdot) \in \{0, 1\}$ is a nearest neighboring indicator function. Suppose $d(z, x_i)$ is the distance between samples z and x_i, if $d(z, x_i)$ is one of the top k minimum distances, $\eta^k(z, x_i) = 1$. Otherwise, $\eta^k(z, x_i) = 0$.

Fault detection task is a bit different from classification problem, where sometimes there are only normal samples in the training sets. Besides, there exist situations that faulty samples are dispersed, wihch traditional KNN is not able to handle. In [35], the authors modified the KNN rule for fault detection problem, which only needs normal training samples, and the KNN rule is expressed as

$$h^{\mathrm{KNN}}(z) = \mathrm{sign}\Big(\varepsilon - \sum_{i=1}^{n} \eta^k(z, x_i)\, d^2(z, x_i)\Big), \tag{2}$$

where ε denotes a threshold. If $h^{\mathrm{KNN}}(z) = 1$, the sample is fault free; otherwise faulty. More specifically, if the sum of square distances from the sample z to its k nearest training samples is smaller than the threshold ε, z is regarded as normal; otherwise, z is faulty. In other words, distance from an abnormal sample to the k nearest neighboring normal training samples must be greater than that from a normal sample.

In this rule, ε is obtained by calculating the distribution of normal training samples square distances to their k nearest neighboring training samples [35]. If v_j is used to represent the sum of square distances from the normal sample x_j to its k nearest normal training samples

$$v_j = \sum_{i=1, i\neq j}^{n} \eta^k(x_j, x_i)\, d^2(x_j, x_i), \tag{3}$$

ε can be defined as

$$\varepsilon = \theta \max_j (v_j), \tag{4}$$

where θ is a constant parameter which is determined by specific application.

III. THE MTS SIMILARITY MEASUREMENT USING MAHALANOBIS DISTANCE BASED DYNAMIC TIME WARPING

The KNN rule is based on similarity measurement of samples. In this work, MTS pieces are extracted as the training and predicting samples. How to measure the similarity of MTS pieces is the core problem in the proposed fault detection framework. In this section, we will introduce the construction of MTS pieces and how to measure the similarity of MTS pieces.

A. THE CONSTRUCTION OF MTS PIECES

Define X as a measurement signal

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}, \tag{5}$$

where x_n represents the m-dimensional observation vector in the training set, and n stands for the length of the measurement signal. Traditional fault detection methods will treat X as a dataset with n static multivariate feature vectors, while the timing sequence and the connection of these multivariate data will not be considered. In contrast, in DPCA algorithm [34], the authors took account of the serial correlation of these multivariate data, where the original measurement signals were broken into several short multivariate time series in the following manner,

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_l \\ x_2 & x_3 & \cdots & x_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-l+1} & x_{n-l+2} & \cdots & x_n \end{bmatrix}, \tag{6}$$

and the length of these MTS pieces is l. Notice that the timing sequence of the data and the dynamic performance can be reserved if using these MTS pieces. Therefore, in our work, this method is used to cut the measurement signals into $n - l + 1$ MTS pieces.

B. MAHALANOBIS DISTANCE BASED DYNAMIC TIME WARPING

In this subsection, the similarity measurement of MTS samples Z and X_i is discussed, where

$$Z = \begin{bmatrix} z_1(1) & z_2(1) & \cdots & z_l(1) \\ z_1(2) & z_2(2) & \cdots & z_l(2) \\ \vdots & \vdots & \ddots & \vdots \\ z_1(m) & z_2(m) & \cdots & z_l(m) \end{bmatrix}, \tag{7}$$

and

$$X_i = \begin{bmatrix} x_{i-l+1}(1) & x_{i-l+2}(1) & \cdots & x_i(1) \\ x_{i-l+1}(2) & x_{i-l+2}(2) & \cdots & x_i(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_{i-l+1}(m) & x_{i-l+2}(m) & \cdots & x_i(m) \end{bmatrix}. \tag{8}$$

In fact, the similarity measurement of the two MTS samples is not easy. On one hand, there might be no one-to-one correspondence between Z and X_i because of different frequencies and phases, so that we should consider the synchronization when measuring different variables of MTS. How to align these two MTS is a challenge. On the other hand, variables play different roles in the system, some of which have strong correlation with the faults, while others may have weak or no relation at all. How to lay stress on the relevant components and decrease the effect of irrelevant dimensions is a big problem as well.

In our previous work [18], traditional dynamic time warping is extended to solve the problem of measuring divergence of two MTS pieces. The optimal warp path W is used to represent the alignment of two MTS samples, which is expressed as

$$W = \begin{pmatrix} w_x(k) \\ w_z(k) \end{pmatrix}, \quad k = 1, 2, \cdots, p \tag{9}$$

where $w_x(k)$ represents a column index from X_i, and $w_z(k)$ represents a column index from Z. p is the length of the warp path W. $(w_x(k), w_z(k))'$ indicates that the $w_x(k)^{th}$ column in X_i corresponds to the $w_z(k)^{th}$ column in Z.

When constructing the warp path W, there are mainly two constraints [38]. The first constraint is that all column indices of the two MTS samples should be found in the warp path W, while the second one is that the warp path W should be continuous and monotonically increasing. Therefore, the starting point and ending point of warp path W is restricted as $W(1) = (1, 1)'$ and $W(p) = (m, n)'$. At the same time, these two constraints also require adjacent points $W(k)$ and $W(k+1)$ to satisfy

$$\begin{cases} w_x(k) \le w_x(k+1) \le w_x(k) + 1 \\ w_z(k) \le w_z(k+1) \le w_z(k) + 1. \end{cases} \tag{10}$$

Therefore, once $W(k)$ is given, there are only three choices for $W(k+1)$, that is $(w_x(k), w_z(k+1))'$, $(w_x(k+1), w_z(k))'$, and $(w_x(k+1), w_z(k+1))'$. Meanwhile, the length of W satisfy that $p \in [l, 2l]$.

Using the optimal warp path W, the MTS X_i and Z can be extended to two new MTS $(\bar{X}_i)_{m\times p}$ and $(\bar{Z})_{m\times p}$, expressed as

$$\begin{cases} \bar{X}_i^k = X_i^{(w_x(k))} \\ \bar{Z}^k = Z^{(w_z(k))} \end{cases} \quad k = 1, 2, \cdots, p \tag{11}$$

where the superscript represents the column index of the MTS. The dynamic time warping (DTW) distance between MTS X_i and Z can be represented by the distance between these two extended MTS $\bar{X}_i$ and $\bar{Z}$, i.e.

$$DTW(X_i, Z) = \sum_{k=1}^{p} D\left(X_i^{w_x(k)}, Z^{w_z(k)}\right) = \sum_{k=1}^{p} D\left(\bar{X}_i^k, \bar{Z}^k\right), \tag{12}$$

where $D(,)$ is a local distance function, and $D\left(X_i^u, Z^v\right)$ represents local distance between the u^{th} column of X_i and the v^{th} column of Z. If $D(,)$ is chosen as the Euclidean distance, it can be expressed as

$$D\left(X_i^u, Z^v\right) = \left(X_i^u - Z^v\right)^T \left(X_i^u - Z^v\right). \tag{13}$$

However, the Euclidean distance is not appropriate to measure the distances among local vectors in many applications. One deficiency of Euclidean distance is that it assigns the same weight to each variable, which is not practical in many situations. First, each variable may have different measurement units in the collection process. Second, some of them are intrinsic and important while others are mixed with noise and outliers. Using the same weight will degrade the classification results. Third, some variables may have a coupling relationship. Noise and outliers in one variable would affect several other variables. Hence, different variables will play different roles in determining the categories of instances. Therefore, the Euclidean distance is not able measure the local distance accurately in these situations. A feasible strategy is to use the Mahalanobis distance function to measure the local distance of vectors in MTS. The Mahalanobis distance between X_i^u and Z^v is expressed as

$$D_M\left(X_i^u, Z^v\right) = \left(X_i^u - Z^v\right)^T M \left(X_i^u - Z^v\right). \tag{14}$$

Here, M is a symmetric positive semi-definite matrix, namely the Mahalanobis matrix. If using Mahalanobis distance as local distance function, the corresponding dynamic time warping is named as the Mahalanobis distance based dynamic time warping (MDDTW).

The MDDTW algorithm can be summarized as the following steps [38], [39]. First of all, a cost distance matrix $Dist_M(u, v)$, where $u = 1, 2, \cdots, l$, $v = 1, 2, \cdots, l$, is constructed. In this matrix, each element $Dist_M(u, v)$ represents the minimum warp distance of sub MTS X_i of length i and sub MTS Z of length j. The corresponding path is named as W_{ij}. Then, as mentioned above, the warp path W_{uv} includes $(u, v)'$ and one of the following choice: $(u-1, v)'$, $(u, v-1)'$ or $(u-1, v-1)'$, which can construct the relationship between $Dist_M(u, v)$ and $Dist_M(u-1, v-1)$, $Dist_M(u-1, v)$ or $Dist_M(uv-1)$. Because $Dist_M(u, v)$ represents the minimum warp distance, thus the relationship is expressed as

$$Dist_M(u, v) = D_M\left(X_i^u, Z^v\right) + \min \begin{cases} Dist_M(u-1, v-1) \\ Dist_M(u-1, v) \\ Dist_M(u, v-1), \end{cases}$$

where $Dist_M(1, 1) = D_M\left(X_i^1, Z^1\right)$. After computing all the elements in the cost distance matrix $Dist_M(i, j)$, $Dist_M(l, l)$ equals to the minimum warp distance $DTW_M(X_i, Z)$ between MTS X_i and Z, which could be rewritten as

$$\begin{aligned} DTW_M(X_i, Z) &= \sum_{k=1}^{p} D_M\left(\bar{X}_i^k, \bar{Z}^k\right) \\ &= \sum_{k=1}^{p} \left(\bar{X}_i^k - \bar{Z}^k\right)^T M \left(\bar{X}_i^k - \bar{Z}^k\right) \\ &= trace\left(P^T M P\right), \end{aligned} \tag{15}$$

where $P_{m\times p} = (\bar{X}_i)_{m\times p} - (\bar{Z})_{m\times p}$. The corresponding warp path W is the optimal warp path.

There are several advantages when using MDDTW algorithm to measure the similarity of MTS pieces. First of all, the variables of MTS stretch or shrink along time axis integrally instead of independently. This will not break the relationship among variables. Besides, a good Mahalanobis distance will rebuild an accurate relationship among variables. The noise and outliers in some variables will be suppressed when measuring the divergence of MTS. Furthermore, the MDDTW measure can be expressed as a very simple form like 15, which is benefitial for the development of the corresponding metric learning algorithm.

IV. ONE-CLASS LogDet DIVERGENCE BASED METRIC LEARNING ALGORITHM FOR FAULT DETECTION

As mentioned above, the proposed MDDTW has its own advantages when comparing MTS pieces. In this approach, the Mahalanobis distance function plays a key role as it can reveal the relationship between variables and categories of instances accurately. Therefore, it is important to learn an appropriate Mahalanobis matrix. In this section, we would like to introduce a metric learning algorithm for fault detection.

A. CONSTRAINTS OF METRIC LEARNING

Our previous work [18] introduces a LogDet divergence based metric learning model for MTS classification. However, the fault detection problem is a little different. In MTS classification problem, MTS with the same label always cluster together, while in the fault detection problem, only fault-free MTS cluster together. The faulty MTS samples with same label have dispersed distributions.

Define that the MTS pieces from normal measurement signals are positive samples, denoted by X_i^+ and X_j^+. while those from faulty signals are called negative samples, denoted by X_k^-. Then, the goal of metric learning in fault detection problem is to learn a Mahalanobis matrix M and corresponding threshold ε that any normal MTS X_j^+ satisfies

$$\sum_{i=1}^{n} \eta^k \left(X_j^+, X_i^+\right) DTW_M \left(X_j^+, X_i^+\right) \le \varepsilon. \tag{16}$$

At the same time, using the same M and ε, any fault MTS X_k^- satisfies

$$\sum_{i=1}^{n} \eta^k \left(X_k^-, X_i^+\right) DTW_M \left(X_k^-, X_i^+\right) > \varepsilon. \tag{17}$$

Approximatively, the following relationship can be used to represent the constraints in fault detection problem,

$$DTW_M \left(X_i^+, X_j^+\right) - DTW_M \left(X_i^+, X_k^-\right) < -\rho, \tag{18}$$

where $\rho > 0$ represents the target margin, and the corresponding $\left\{X_i^+, X_j^+, X_k^-\right\}$ is called as triplet constraints.

B. LogDet DIVERGENCE BASED METRIC LEARNING MODEL

The above mentioned MTS construction method always produces a huge amount of MTS pieces, so it is necessary to guarantee that the metric learning process is scalable with respect to the size of the training samples. Thus, we adopt an online metric learning framework [41], [42] to learn the Mahalanobis matrix. In this model, the Mahalanobis matrix changes gradually as the model trains one triplet constraint at a time.

Define M_t as a known matrix representing the current obtained Mahalanobis matrix at time step t. When the metric learning model receives a triplet constraint $\left(X_i^+, X_j^+, X_k^-\right)$, there will be no loss if the triplet constraint satisfies the relationship in the inequation (18). Otherwise, the current M_t should be updated to a better one to reduce the following loss function,

$$l(M) = \rho + DTW_M \left(X_i^+, X_j^+\right) - DTW_M \left(X_i^+, X_k^-\right). \tag{19}$$

At the next time step $t+1$, the M_{t+1} will be updated to satisfy $M_{t+1} = \arg\min_M l(M)$. When the total loss function $L(M) = \sum_t \ell(M_t)$ reaches its minimum, the corresponding M will be the objective optimal Mahalanobis matrix.

Because the optimization of M is decomposed into multiple steps, it is important to make sure the learning process is stable. Therefore, the model needs a regularization term to guarantee that the Mahalanobis matrix changes gradually and stably. In this work, we use the logDet divergence [43] to measure the divergence between M and M_t, expressed as

$$D_{ld}(M, M_t) = tr\left(M M_t^{-1}\right) - \log\left(\det\left(M M_t^{-1}\right)\right) - n. \tag{20}$$

where n is the dimension of M. The logDet divergence based metric learning model for MTS is to solve the following iterative minimization problem,

$$M_{t+1} = \arg\min_{M \succ 0} D_{ld}(M, M_t) + \eta_t \ell(M) \tag{21}$$

where $\eta_t > 0$ is a regularization parameter. The η_t is used for the balance of the regularization function $D_{ld}(M, M_t)$ and loss function $\ell(M)$. The $D_{ld}(M, M_t) + \eta_t \ell(M)$ reaches its extremum value when its gradient is zero. Since its second order derivative of $D_{ld}(M, M_t) + \eta_t \ell(M)$ is $M^{-2} \ge 0$, the extremum value is the global minimum. We obtain the following equation by setting the gradient of (21) to be zero with respect to M,

$$M_{t+1} = \left(M_t^{-1} + \eta_t \left(PP^T - QQ^T\right)\right)^{-1}, \tag{22}$$

where $(P_t)_{m\times p} = \left(\bar{X}_i^+\right)_{m\times p} - \left(\bar{X}_j^+\right)_{m\times p}$ and $(Q_t)_{m\times q} = \left(\bar{X}_i^+\right)_{m\times q} - \left(\bar{X}_k^-\right)_{m\times q}$. p and q represent the number of columns which are calculated by using the MDDTW algorithm. To avoid expensive computation of matrix inverse, we apply the Woodbury matrix identity to solve (22). The standard Woodbury matrix identity is

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1} VA^{-1}. \tag{23}$$

However, in the updating equation, there are two items which are the outer product of matrices. To solve this problem, we assume that $\gamma_t = \left(M_t^{-1} + \eta_t P_t P_t^T\right)^{-1}$, and (22) is splited into two standard Woodbury matrix identity equations,

$$\begin{cases} \gamma_t = \left(M_t^{-1} + \eta_t PP^T\right)^{-1} \\ M_{t+1} = \left(\gamma_t^{-1} - \eta_t QQ^T\right)^{-1}. \end{cases} \tag{24}$$

Applying the Woodbury matrix identity, we arrive at an analytical expression for M_{t+1}

$$\begin{cases} \gamma_t = M_t - \eta_t M_t P\left(I+\eta_t P^T M_t P\right)^{-1} P^T M_t \\ M_{t+1} = \gamma_t + \eta_t \gamma_t Q\left(I - \eta_t Q^T \gamma_t Q\right)^{-1} Q^T \gamma_t, \end{cases} \tag{25}$$

where the regularization parameter η_t is chosen as $\eta_t = \alpha\bar{\eta}_t$. α is a constant learning rate parameter which is chosen between 0 and 1, and $\bar{\eta}_t$ is evaluated by solving the following linear matrix inequalities

$$\begin{cases} \bar{\eta}_t \left(PP^T - QQ^T\right) + M_t^{-1} \geq 0 \\ \bar{\eta}_t \geq 0 \end{cases}. \tag{26}$$

Thanks to the MDDTW, the Mahalanobis metric can be trained uniformly using MTS with various lengths and phases, and the convergence and boundedness have been proved in [44]. However, there is a weak point of the proposed method that the leaning process is time consuming. The computational complexity is $O\left(Nd^2l^2\right)$, where N is the number of triplets. To solve this problem, our strategy is to select partial triplet constraints instead of all triplets in the metric learning process.

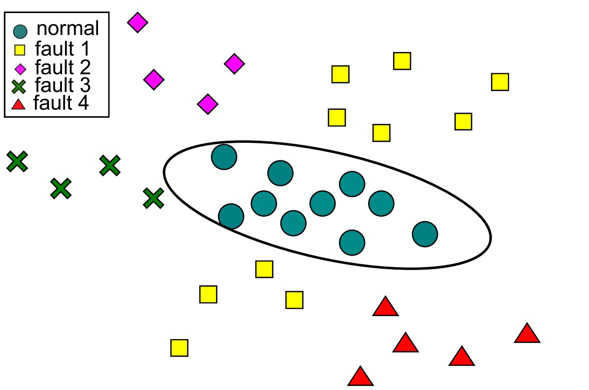


FIGURE 2. **One possible data distribution in fault detection problems.**

C. DYNAMIC TRIPLETS BUILDING STRATEGY

In the metric learning process, it is impossible to utilize all triplets because the total quantity of triplets is cubic in the volume of training MTS. It is important to select the most useful triplets and remove redundant ones. In our previous work [44], a dynamic triplets building strategy was proposed. The metric learning process is broken into several cycles. According to the Mahalanobis distance matrix of training data pairs, the most useful triplets are chosen in each cycle. In fault detection problem, the data distribution is different from the normal classification problem. Fig. 2 illustrates one possible data distribution of fault detection problem. The normal data have a concentrated distribution while fault data have dispersed distribution, and divergence of the fault data with the same class may be even larger than that between fault and normal data. Therefore, in this paper, we develop a dynamic triplets building strategy for fault detection problem.

In this strategy, the metric learning process is divided into several cycles. The category of normal MTS is named as C_N while the category of all other fault MTS is named as C_F. In the first step, the triplets at the boundaries of these two categories are chosen in the initial cycle. For each MTS X_i^+, instances X_j^+ which has the largest Euclidean distance in C_N and instances X_k^- which has the nearest Euclidean distance in C_F are selected, and these $\{(i, j, k)\}$ are picked to build the training triplets [45]. However, these selected triplet constraints $\{(i, j, k)\}$ are based on Euclidean distance. Then, the triplets set should be updated in the following learning process. Using these triplet constraints, a Mahalanobis matrix M_t can be obtained using the LogDet divergence based metric learning algorithm after a metric learning cycle. Then, we use the MDDTW method to measure the performance of M_t on the training MTS. After that, new triplets in the next cycle are selected based on the measurement.

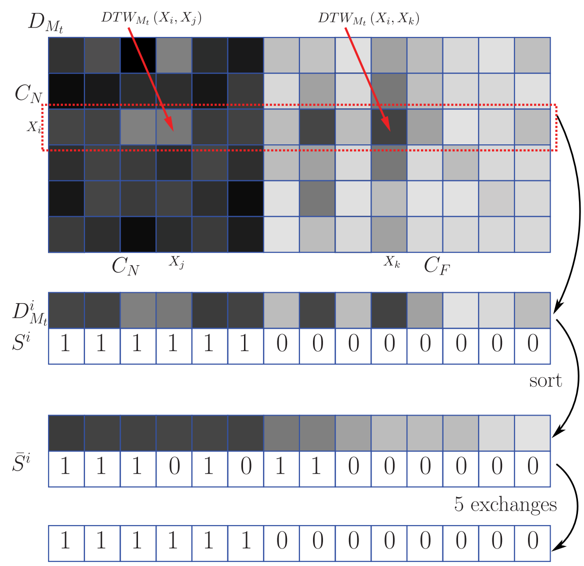


FIGURE 3. **The framework of the proposed triplets building strategy.**

The key problem in this method is how to measure the performance of M_t on the training MTS set. Two matrices are defined here. The first matrix is the MDDTW distance matrix $D_{M_t} = \left\{DTW_{M_t}\left(X_i, X_j\right)\right\}$, which represents the MDDTW distance between positive training MTS and all training MTS (including positive and negative samples). The second one is the similarity matrix $S = \left\{s\left(X_i, X_j\right)\right\}$, which gives the relationship information of sample pairs. If X_j is a positive MTS, $s\left(X_i, X_j\right) = 1$, otherwise, $s\left(X_i, X_j\right) = 0$. Fig. 3 illustrates the framework of the proposed triplets building strategy. The top section of Fig. 3 shows the MDDTW distance matrix $D_{M_t} = \left\{DTW_{M_t}\left(X_i, X_j\right)\right\}$. The instances in training set have been sorted according to their categories. The positive samples in C_N are sorted in front of negative samples in C_F. In fact, there are two main blocks in the MDDTW distance matrix. The first one is called intra-class block representing the MDDTW distance among instances in the C_N while the other block (inter-class block) represents the MDDTW distance between instances in C_N and C_F. The deep color means the element value is small while light color represents a large value. The ideal situation is that the MDDTW distances in intra-class block should be all smaller than that in inter-class block. However, some points $DTW_{M_t}\left(X_i^+, X_k^-\right)$ in inter-class block has a deeper color than the points $DTW_{M_t}\left(X_i^+, X_j^+\right)$ in intra-class block. That means the obtained Mahalanobis distance M_t has't achieved the best performance, and these triplets $\{(i, j, k)\}$ should be picked as triplet constraints in the next metric learning cycle.

We extract the i^{th} row in D_{M_t} which is expressed as $D^i_{M_t} = \left\{DTW_{M_t}\left(X_i, X_j\right)\right\}$ and the corresponding $S^i = \{s\left(x_i, x_j\right)\}$, where $j = 1, 2, \cdots, n$. If the vector $DTW^i_{M_t}$ is sorted in an ascending order, the corresponding S^i will be changed to $\bar{S}^i$. The ideal case is that the $\bar{S}^i$ should be in a descending order. However, there will be 0 elements in the $\bar{S}^i$ is in front of 1, which means that the MDDTW distance between several negative and positive pairs is smaller than that of positive pairs using the current M_t. If the elements are to be reordered in the $\bar{S}^i$, we should exchange the disordered 0 and 1 elements for five times, and we name this exchange count as the disorder Θ_i. The normalized disorder $\theta_i = \Theta_i / \sum_i \Theta_i$ is used to determine the proportion of triplet constraints which contain x_i. When building triplets including x_i, the method is similar to the method in [45]. The main difference is that Mahalanobis matrix M_t is used instead of Euclidean distance. It worth noting that the total disorder $\tau = \sum_i \Theta_i$ can be used to judge whether the algorithm is converged.

In this dynamic triplets building strategy, a feedback from the current Mahalanobis matrix to triplet constraints is constructed, where the negative feedback would guarantee the stability. The most useful triplets in each cycle are selected, which is beneficial for efficient and stable metric learning process. In this process, the main time consumption is the calculation of D_{M_t}, and the time complexity of the triplets building process is about $O\left(l^2\left(n_N^2 + n_N n_F\right)\right)$.

V. EXPERIMENTS AND RESULTS

In this work, experiments on the TE process [31] are conducted to illustrate the performance of the proposed fault detection framework. All experiments are tested in MATLAB 2016a, and all tests are implemented on a computer with Intel(R) Core(TM) i5-4430, 3.00GHz CPU, 8G RAM, and Windows 10 operating system. The relevant source codes can be found on the MATLAB central.[1] Besides, two general indices are adopted to evaluate the fault diagnosis performance, i.e. fault detection rate (FDR) and false alarm rate (FAR). Assume that tp, tn, fp and fn represent correct fault detection results, correct normal signal detection results, false fault alarm results, and false fault missing results, respectively. FDR and FAR can be expressed as

$$\begin{cases} FDR = \dfrac{tp}{tp + fn} \times 100\% \\ FAR = \dfrac{fp}{fp + tn} \times 100\%. \end{cases} \tag{27}$$

The TE process is a realistic simulation program of a chemical plant which has been widely studied as a benchmark in many fault detection methods. In this work, the data sets of the TE process are downloaded from the website.[2] In this database, there are 21 different kinds of faults, named as IDV(1), IDV(2), $\cdots$, IDV(21) [32], [33]. IDV(1-20) are process faults while IDV(21) is an additional value fault. There are 22 training sets and 22 testing sets in the database, including 21 faulty data files and one normal data file. Each training data file contains 480 rows which record 52 variables for 24 operation hours. It is worth noting that only 11 manipulated variables and 22 process variables are available in the fault detection process. The rest 19 analysis variables are obtained after operation of the system. Therefore, we only consider the available 33 variables in the experiments. The data in each testing data file are collected via 48 hour plant operation time, in which the fault occurs at the beginning of the 8^{th} operation hour. In other word, each testing data file contains 960 rows, where the first 160 points are normal data and the rest 800 points are abnormal data.

In the first experiment, a Mahalanobis distance function is learnt based on the proposed approach and used for fault detection. The length of the MTS pieces is chosen as $l = 16$. The MTS constructed from normal data file is used as positive samples, and all the 21 faulty data files construct the negative samples. The threshold is chosen as $\varepsilon = 1.25 \max(v)$. The test results of the proposed MDDTW based fault detection method are given in the last column of Table 1. Meanwhile, we also conduct the same experiment using Euclidean distance based dynamic time warping (EDDTW), and the detected results are also presented in Table 1. Besides, this table also records the results of other classical fault detection methods, including PCA, DPCA, ICA, MICA, FDA, PLS, TPLS, MPLS and SAP, the results of which have been reported by literature [33].

In [33], 21 fault signals of TE process are classified as three categories. The first type consists of IDV(1-2), IDV(4-8), IDV(12-14) and IDV(17-18). These faults can be easily detected, and almost all methods have high FDRs as well as low FARs in fault detection results. The faults in the second category, including IDV(10-11), IDV(16), IDV(19) and IDV(20-21) are relatively not easy to be detected. The fault detection performances with different methods on these faults are in various precision levels. In the third category, all methods produce bad results because these

[1] https://ww2.mathworks.cn/matlabcentral/fileexchange/67582-mahalanobis-distance-based-dynamic-time-warping-for-fault-detection

[2] http://brahms.scs.uiuc.edu

TABLE 1. Results comparison for different fault detection methods on TE data sets given in [32].

Fault (FDR (%))	PCA	DPCA	ICA	MICA	FDA	PLS	TPLS	MPLS	SAP	EDDTW	MDDTW
IDV(1)	99.88	99.88	**100**	99.88	**100**	99.88	99.88	**100**	99.63	99.50	99.50
IDV(2)	98.75	**99.38**	98.25	98.25	98.75	98.63	98.88	98.88	97.88	98.50	98.25
IDV(4)	**100**	**100**	**100**	87.63	**100**	99.50	**100**	**100**	99.88	99.50	**100**
IDV(5)	33.63	43.25	**100**	**100**	**100**	33.63	**100**	**100**	**100**	30.25	99.88
IDV(6)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	99.13	**100**
IDV(7)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	99.88	**100**	**100**
IDV(8)	98.00	98.00	98.25	97.63	98.13	97.88	98.50	**98.63**	95.88	97.38	97.25
IDV(12)	99.13	99.25	**99.88**	**99.88**	99.75	99.25	99.63	**99.88**	**99.88**	99.63	99.75
IDV(13)	95.38	95.38	95.25	95.00	95.63	95.25	**96.13**	95.50	94.88	94.38	94.75
IDV(14)	**100**	**100**	**100**	99.88	**100**	**100**	**100**	**100**	97.63	99.50	99.88
IDV(17)	95.25	**97.25**	96.88	93.00	96.63	94.25	96.00	97.13	97.13	81.38	**97.25**
IDV(18)	90.50	90.88	90.50	89.75	90.75	90.75	**91.88**	91.25	91.00	89.63	90.13
IDV(10)	60.50	72.00	89.25	85.88	87.13	82.63	91.00	91.13	**95.50**	47.75	95.13
IDV(11)	78.88	91.50	78.88	61.63	73.38	78.63	86.13	83.25	84.75	90.38	**98.50**
IDV(16)	55.25	67.38	92.38	83.38	83.25	68.38	90.75	94.28	94.88	92.50	**99.13**
IDV(19)	41.13	87.25	92.38	80.25	87.88	26.00	82.88	94.25	88.50	65.50	**99.88**
IDV(20)	63.38	73.75	91.38	86.00	81.88	62.75	78.38	91.50	83.75	86.88	**91.75**
IDV(21)	52.13	61.00	56.38	70.75	52.75	59.88	66.38	**72.75**	38.63	45.38	67.25
IDV(3)	12.88	12.25	4.50	14.25	7.00	14.25	**24.25**	18.75	6.38	0.25	3.13
IDV(9)	8.38	12.88	4.75	8.88	6.25	14.50	**23.50**	12.13	0.88	0	0.75
IDV(15)	14.13	19.75	7.75	10.75	12.63	23.00	**29.88**	23.25	29.50	3.25	12.88
Fault free (FAR (%))	PCA	DPCA	ICA	MICA	FDA	PLS	TPLS	MPLS	SAP	EDDTW	MDDTW
IDV(0)	6.13	10.13	2.75	1.63	6.38	10	19.62	10.75	1.50	1.27	**1.16**

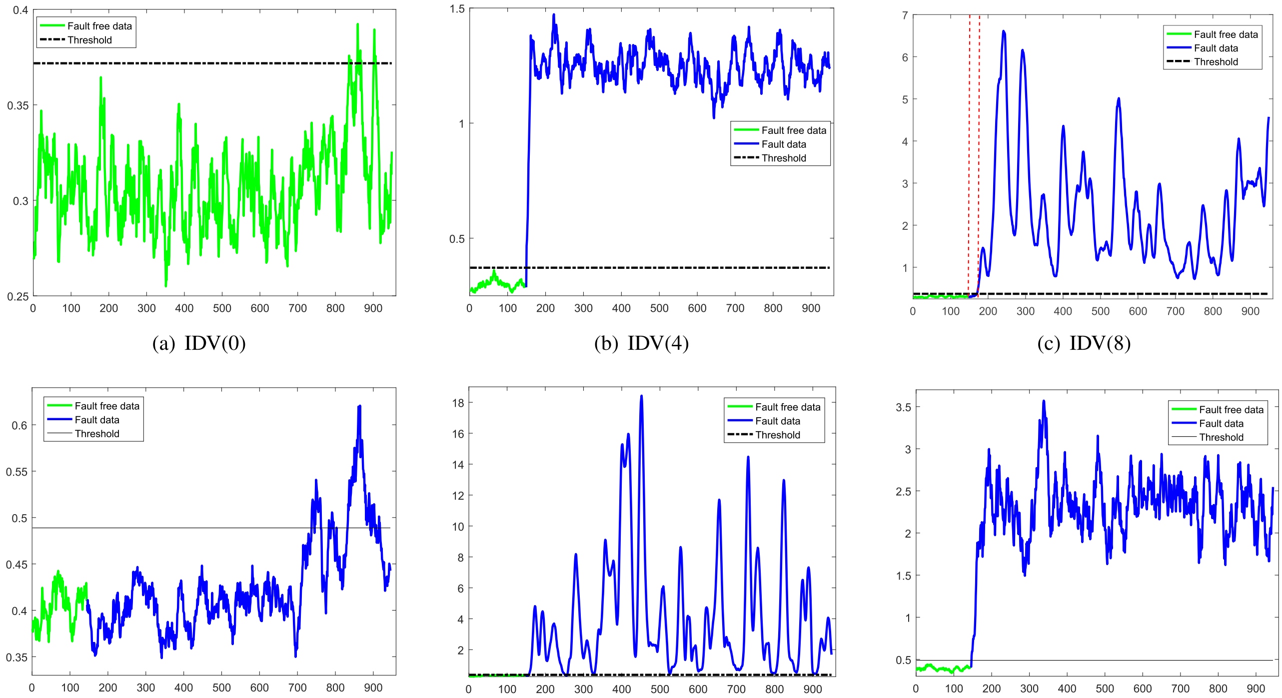
faults are very difficult to be detected. Besides, the IDV(0) denotes fault free signal which is used to evaluate the FARs.

It can be seen from the results that the proposed method has good performance on faulty signals in the first and second categories. For the first category, the proposed method has similar performance as other methods. However, it achieves 100% FDR only on IDV(4), IDV(6) and IDV(7), e.g. Fig. 4(b) illustrates the fault detection result for IDV(4). For other fault signals in the first category, the FDRs can hardly reach 100%. This is because the proposed method chooses MTS to judge the status of the system. There are some lags as MTS occupy a period of time. Therefore, the faults would be detected a little later after the fault occurs as shown in Fig. 4(c). For the second category, the proposed method greatly improves the fault detection performance, especially on IDV(11), IDV(16), IDV(19) and IDV(20). As mentioned above, these 4 faulty signals have a common feature, i.e. when fault occurs, the system static index jumps between normal and fault status. The signals in normal status will be judged as positive while signals in fault status will be labelled as negative if using static fault detection methods. Therefore, these methods all produce low FDRs. The proposed method uses MTS as detecting samples, and the jumping process can be recorded. If there exists jumping process in the testing MTS, there will be large MDDTW divergence between the testing MTS and normal MTS. Thus the proposed method can easily achieve good FDRs in this kind of situations as shown in Fig. 4(e) and Fig. 4(f). This feature is also the main advantage of the proposed fault detection method. For the third category, these 3 faulty signals can hardly be detected with our approach, similar to other fault detection methods, e.g. Fig. 4(d) gives the detection results on the IDV(15). The last row of Table 1 illustrates the FARs of faults on the fault free signal IDV(0), where the proposed method achieves the lowest FAR on IDV(0), which is also shown in Fig. 4(a).

TABLE 2. The relationship between FDR(%) and MTS length.

Fault	4	8	12	16	20	24
IDV(8)	97.63	97.50	97.50	97.25	97.25	97.25
IDV(13)	95.00	95.00	95.25	94.75	94.75	94.63
IDV(11)	85.50	93.25	98.25	98.50	99.13	99.25
IDV(16)	95.38	97.63	98.50	99.13	99.13	99.25
IDV(21)	57.00	57.63	67.13	67.25	70.50	71.00

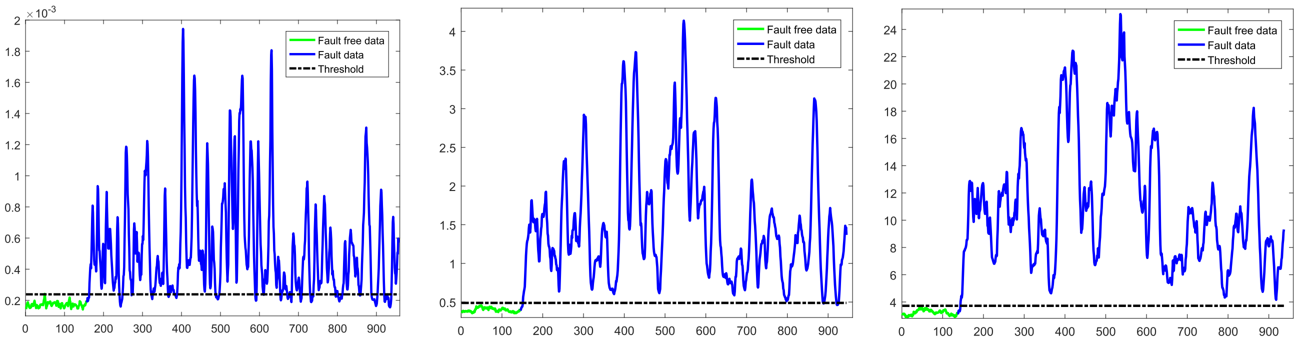
In the second experiment, we explore the relationship between FDR and the length of MTS. MTS with different data length are constructed ranging from 4 to 24. Then the corresponding Mahalanobis distance functions are learnt and applied in the fault detection process. The FDR with different MTS lengths on IDV(8), IDV(11), IDV(13), IDV(16) and IDV(21) are illustrated in Table 2. In the experiment, all the thresholds are chosen as $\varepsilon = 1.25 \max(v)$. It can be observed that FDRs for the first and second categories have opposite trend when MTS length grows. FDR of faulty signals in the first category decreases slightly while it grows in the second category. For example, FDR of IDV(8) and IDV(13) decreases by 0.38% when the MTS length increases from 4 to 24. For IDV(11), IDV(16) and IDV(20), FDR increases rapidly from 4 to 12, then grows slowly from 12 to 24. The reason for this relationship between FDR and MTS length can be explained as follows. On one hand, the fault in the first category can be easily detected since the static data distributions of normal and faulty signals have significant difference. When increasing the MTS length, the lag effect will be stronger, so the FDR will decrease. On the other hand, the main feature of faulty signals in the second category is the



(a) IDV(0) (b) IDV(4) (c) IDV(8)

(d) IDV(15) (e) IDV(16) (f) IDV(19)

FIGURE 4. Process monitoring using the proposed method.



(a) MTS length=4 (b) MTS length=16 (c) MTS length=24

FIGURE 5. Process monitoring on IDV(11) using the proposed method with different MTS length.

jumping characteristics, and longer MTS can record this process more completely, which is beneficial for fault detection. Therefore, the FDR grows rapidly. When the length of MTS is enough to recover a whole jumping process, the FDR will slowly increase. Fig. 5 illustrates the fault detection results on IDV(11) using the proposed method with different MTS lengths. It can been noticed that longer MTS can produce smoother fault detection results, and the divergence between normal and faulty data becomes clearer.

VI. CONCLUSIONS

In this paper, a novel data-driven fault detection framework is proposed to deal with the process control and fault diagnosis problems in industrial applications. In this framework, MTS pieces are used to represent the dynamic features of the measurement signals, and a novel similarity metric called Mahalanobis distance based dynamic time warping for MTS pieces is proposed. The MDDTW can stretch or shrink the variables of MTS pieces along time axis integrally and find the optimal warp path to achieve the one-to-one correspondence of MTS pieces. Meanwhile, the MDDTW uses Mahalanobis distance to build an accurate relationship among variables, which are beneficial for suppressing noises and outliers in some variables when comparing MTS samples. Besides, this work also puts forward a one-class metric learning algorithm for the fault detection method. The proposed metric learning framework uses a LogDet divergence model as well as a one-class dynamic triplets building strategy, which can learn the Mahalanobis distance accurately and robustly. The k-nearest neighbor rule is chosen as the fault detection criterion, which needs no assumption on the probability distribution of normal and faulty measurement signals. Experimental results on the famous TE process demonstrate that the proposed fault detection framework has improved performance compared with classical methods, especially when dealing with the faulty signals with jumping characteristics. The weak points of the algorithm are the huge amount of computations in the off-line training process and delay effect in the on-line detection stage. These are the two problems to be solved in future works.

ACKNOWLEDGMENT

The authors also appreciate the kind suggestions on this work from Dr. Jiangyuan Mei in Midea Corporate Research Center. The computational resources in the HPC Center of Zhejiang University (Zhoushan Campus) are also acknowledged.

REFERENCES

[1] A. Ghosh, S. Qin, J. Lee, and G.-N. Wang, "FBMTP: An automated fault and behavioral anomaly detection and isolation tool for PLC-controlled manufacturing systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 12, pp. 3397–3417, Dec. 2017.

[2] J. Neuzil, O. Kreibich, and R. Smid, "A distributed fault detection system based on IWSN for machine condition monitoring," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1118–1123, May 2014.

[3] S. Yin, H. Gao, J. Qiu, and O. Kaynak, "Fault detection for nonlinear process with deterministic disturbances: A just-in-time learning based data driven method," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3649–3657, Nov. 2017.

[4] P. Qian, X. Ma, D. Zhang, and J. Wang, "Data-driven condition monitoring approaches to improving power output of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6012–6020, Aug. 2019.

[5] M. Schwabacher, "A survey of data-driven prognostics," in *Proc. AIAA Infotech Aerosp. Conf.*, 2005, p. 7005.

[6] P. Zhang, H. Ye, S. X. Ding, G. Z. Wang, and D. H. Zhou, "On the relationship between parity space and approaches to fault detection," *Syst. Control Lett.*, vol. 55, no. 2, pp. 94–100, Feb. 2006.

[7] W. Chen and M. Saif, "Observer-based fault diagnosis of satellite systems subject to time-varying thruster faults," *J. Dyn. Syst., Meas., Control*, vol. 129, no. 3, pp. 352–356, May 2007.

[8] Z. Gao, H. Saxen, and C. Gao, "Guest editorial: Special section on data-driven approaches for complex industrial systems," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2210–2212, Nov. 2013.

[9] S. Yin, G. Wang, and H. R. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, vol. 24, no. 4, pp. 298–306, Jun. 2014.

[10] L. Li, S. X. Ding, J. Qiu, Y. Yang, and D. Xu, "Fuzzy observer-based fault detection design approach for nonlinear processes," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 8, pp. 1941–1952, Aug. 2017.

[11] 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.

[12] W. Sun, X. Wang, and C. Zhang, "A model-free control strategy for vehicle lateral stability with adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, early access, Dec. 12, 2019, doi: 10.1109/TIE.2019.2958308.

[13] J. Zhang, H. Chen, S. Chen, and X. Hong, "An improved mixture of probabilistic PCA for nonlinear data-driven process monitoring," *IEEE Trans. Cybern.*, vol. 49, no. 1, pp. 198–210, Jan. 2017.

[14] S. X. Ding, P. Zhang, S. Yin, and E. L. Ding, "An integrated design framework of fault-tolerant wireless networked control systems for industrial automatic control applications," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 462–471, Feb. 2013.

[15] E. Alizadeh, N. Meskin, and K. Khorasani, "A negative selection immune system inspired methodology for fault diagnosis of wind turbines," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3799–3813, Nov. 2017.

[16] Z. Chang, W. Chen, Y. Gu, and H. Xu, "Mahalanobis-taguchi system for symbolic interval data based on kernel mahalanobis distance," *IEEE Access*, vol. 8, pp. 20428–20438, 2020.

[17] L. H. Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometric Intell. Lab. Syst.*, vol. 50, no. 2, pp. 243–252, Mar. 2000.

[18] J. Mei, M. Liu, Y.-F. Wang, and H. Gao, "Learning a mahalanobis distance-based dynamic time warping measure for multivariate time series classification," *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1363–1374, Jun. 2016.

[19] M. Misra, H. H. Yue, S. J. Qin, and C. Ling, "Multivariate process monitoring and fault diagnosis by multi-scale PCA," *Comput. Chem. Eng.*, vol. 26, no. 9, pp. 1281–1293, Sep. 2002.

[20] Y. Zhang, H. Zhou, S. J. Qin, and T. Chai, "Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 3–10, Feb. 2010.

[21] M. R. Maurya, R. Rengaswamy, and V. Venkatasubramanian, "Fault diagnosis using dynamic trend analysis: A review and recent developments," *Eng. Appl. Artif. Intell.*, vol. 20, no. 2, pp. 133–146, Mar. 2007.

[22] S. Yin, S. X. Ding, P. Zhang, A. Hagahni, and A. Naik, "Study on modifications of PLS approach for process monitoring," *IFAC Proc. Volumes*, vol. 44, no. 1, pp. 12389–12394, Jan. 2011.

[23] J. E. Jackson and G. S. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, no. 3, pp. 341–349, Aug. 1979.

[24] N. D. Tracy, J. C. Young, and R. L. Mason, "Multivariate control charts for individual observations," *J. Qual. Technol.*, vol. 24, no. 2, pp. 589–590, 1992.

[25] Q. P. He, S. J. Qin, and J. Wang, "A new fault diagnosis method using fault directions in Fisher discriminant analysis," *AIChE J.*, vol. 51, no. 2, pp. 555–571, 2005.

[26] S. Ding, P. Zhang, E. Ding, A. Naik, P. Deng, and W. Gui, "On the application of PCA technique to fault diagnosis," *Tsinghua Sci. Technol.*, vol. 15, no. 2, pp. 138–144, Apr. 2010.

[27] G. Li, S. J. Qin, and D. Zhou, "Geometric properties of partial least squares for process monitoring," *Automatica*, vol. 46, no. 1, pp. 204–210, Jan. 2010.

[28] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno, "Monitoring independent components for fault detection," *AIChE J.*, vol. 49, no. 4, pp. 969–976, Apr. 2003.

[29] J.-M. Lee, S. J. Qin, and I.-B. Lee, "Fault detection and diagnosis based on modified independent component analysis," *AIChE J.*, vol. 52, no. 10, pp. 3501–3514, 2006.

[30] S. X. Ding, P. Zhang, A. Naik, E. L. Ding, and B. Huang, "Subspace method aided data-driven design of fault detection and isolation systems," *J. Process Control*, vol. 19, no. 9, pp. 1496–1510, Oct. 2009.

[31] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.

[32] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. Springer, 2000.

[33] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process," *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, Oct. 2012.

[34] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometric Intell. Lab. Syst.*, vol. 51, no. 1, pp. 81–93, May 2000.

[35] Q. P. He and J. Wang, "Fault detection using the k-Nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007.

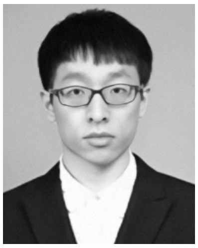
[36] G. Verdier and A. Ferreira, "Adaptive mahalanobis distance and k-nearest neighbor rule for fault detection in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 1, pp. 59–68, Feb. 2011.

[37] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-Nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, Mar. 2016.

[38] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intell. Data Anal.*, vol. 11, no. 5, pp. 561–580, Oct. 2007.

[39] M. Müller, "Dynamic time warping," in *Proc. Inf. Retr. Music Motion*, 2007, pp. 69–84.
[40] G. Al-Naymat, S. Chawla, and J. Taheri, "Sparsedtw: A novel approach to speed up dynamic time warping," in *Proc. 8th Australas. Data Mining Conf.*, vol. 101 2009, pp. 117–127.
[41] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. 24th Int. Conf. Mach. Learn. (ICML)*, 2007, pp. 209–216.
[42] P. Jain, B. Kulis, I. S. Dhillon, and K. Grauman, "Online metric learning and fast similarity search," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 761–768.
[43] B. Kulis, M. Sustik, and I. Dhillon, "Learning low-rank kernel matrices," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 505–512.
[44] J. Mei, M. Liu, H. R. Karimi, and H. Gao, "LogDet divergence-based metric learning with triplet constraints and its applications," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4920–4931, Nov. 2014.
[45] M. Liu and B. C. Vemuri, "A robust and efficient doubly regularized metric learning approach," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012.

YULIN SI (Member, IEEE) was born in China, in 1986. He received the bachelor's and master's degrees in control science and engineering from the Harbin Institute of Technology, in 2009 and 2011, respectively, and the Ph.D. degree in offshore wind energy from the University of Agder, Norway, in 2016. He is currently working as an Assistant Professor with the Ocean College, Zhejiang University. His current research interests include fault diagnosis and fault-tolerant control, and control design for offshore renewable energy systems.



ZHENG CHEN was born in China, in 1994. He received the bachelor's degree in mechanical and electronics engineering from Zhejiang University, in 2017, where he is currently pursuing the Ph.D. degree in ocean engineering and technology. His current research interests include renewable energy platforms design, advanced control, and the condition monitoring of offshore platforms.

JILI SUN was born in China, in 1997. She received the bachelor's degree in ocean engineering and technology from Zhejiang University, in 2019, where she is currently pursuing the Ph.D. degree. Her research interests include the advanced control and condition monitoring of offshore wind systems.

DAHAI ZHANG (Member, IEEE) was born in China, in 1981. He received the Ph.D. degree from Zhejiang University, in 2010. From 2011 to 2013, he was a Postdoctoral Researcher with Zhejiang University and Lancaster University. He is currently a Professor with the Ocean College, Zhejiang University. He has been the PI of various research projects in regards to modeling, design, and measurements of renewable energy electrical machines such as wind turbines, tidal current turbines, and wave energy converters.

PENG QIAN received the B.Eng. degree in electrical engineering and automation and the M.Sc. degree in power electronics and drives from Jiangsu University, China, in 2009 and 2014, respectively, and the Ph.D. degree from Lancaster University, in 2017. From 2017 to 2019, he worked as a Research Fellow of Brunel University, U.K. He is currently the Senior Project Leader of TWI. His research interests include predictive condition monitoring, data driven-based modeling, optimal energy management, and power electronics.