Research Article

Metric Learning Method Aided Data-Driven Design of Fault Detection Systems

Guoyang Yan,[1] Jiangyuan Mei,[1] Shen Yin,[1] and Hamid Reza Karimi[2]

[1] *Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Heilongjiang 150001, China*
[2] *Department of Engineering, Faculty of Engineering and Science, The University of Agder, 4898 Grimstad, Norway*

Correspondence should be addressed to Shen Yin; shen.yin2011@gmail.com

Fault detection is fundamental to many industrial applications. With the development of system complexity, the number of sensors is increasing, which makes traditional fault detection methods lose efficiency. Metric learning is an efficient way to build the relationship between feature vectors with the categories of instances. In this paper, we firstly propose a metric learning-based fault detection framework in fault detection. Meanwhile, a novel feature extraction method based on wavelet transform is used to obtain the feature vector from detection signals. Experiments on Tennessee Eastman (TE) chemical process datasets demonstrate that the proposed method has a better performance when comparing with existing methods, for example, principal component analysis (PCA) and fisher discriminate analysis (FDA).

1. Introduction

Due to the fact that industrial systems are becoming more complex, safety and reliability have become more critical in complicated process design [1–3]. Traditional model-based approaches, which require the process modeled by the first principle or prior knowledge of the process, have become difficult, especially for large-scale processes. With significantly growing automation degrees, a large amount of process data is generated by the sensors and actuators. In this framework, the data-based techniques are proposed and developed rapidly over the past two decades. Data-driven fault diagnosis schemes are based on considerable amounts of historical data, which take sufficient use of the information provided by the historical data instead of complex process model [4, 5]. This framework can simplify the design procedure effectively and ensure safety and reliability in the complicated processes [6]. Many fault diagnosis techniques have been used in the complicated industrial systems [7–9]. In this framework, PCA [10] and FDA [11] are regarded as the most mature and successful methods in real industrial applications.

PCA aims at dimensionality reduction, which captures the data variability in an efficient way. In PCA method, process variables are projected onto two orthogonal subspaces by carrying out the singular value decomposition on the sample covariance matrix. And cumulative percent variance [12] is the standard to determine the number of principal components. To detect the variability information in two orthogonal subspaces, the squared prediction error (SPE) statistic [13] and the T^2 statistic [14] are calculated. PCA is a sophisticated method. However, PCA determines the lower dimensional subspaces without considering the information between the classes. FDA [15] is a linear dimensionality reduction technique. It has advantages over PCA because it takes into consideration the information between different classes of the data. The aim of FDA is to maximize the dispersion between different classes and minimize the dispersion within each class by determining a group of transformation vectors. In FDA method, three matrices are defined to measure dispersion. The problem of determining a set of linear transformation vectors is equal to the problem of solving generalized eigenvalues [16]. However, FDA has difficulty in dealing with online applications. Motivated by

Input: x_k: a given data sets of n points, s: set of pairs of data in same categories, D: set of pairs of data in different categories, H_0: a given matrix, s: a given upper bound, b: a given lower bound, β: the slack variable
Output: H: the target Mahalanobis matrix
(1) $H = H_0$, $\lambda(i, j) = 0$
(2) $m(i, j) = s$ when $(i, j) \in S$, $m(i, j) = b$ when $(i, j) \in D$
(3) Repeat until convergence:
    (3.1) Pick a pairs of data $\left(x_i, x_j\right)_t \in x_k$
    (3.2) $l(i, j)_t = 1$ when $(i, j)_t \in S$, $l(i, j)_t = -1$ when $(i, j)_t \in D$
    (3.3) $g_t = \left(x_i - x_j\right)^T H_t \left(x_i - x_j\right)$
    (3.4) $k_t = l(i, j)_t \cdot \min\left(\lambda(i, j)_t, \dfrac{l(i, j)_t}{2}\left(\dfrac{1}{g_t} - \dfrac{\beta}{m(i, j)_t}\right)\right)$
    (3.5) $\mu_t = \dfrac{k_t}{1 - k_t g_t}$
    (3.6) $m(i, j)_{t+1} = \dfrac{\beta m(i, j)_t}{\beta + k_t m(i, j)_t}$
    (3.7) $\lambda(i, j)_{t+1} = \lambda(i, j)_t - \min\left(\lambda(i, j)_t, \dfrac{l(i, j)_t}{2}\left(\dfrac{1}{g_t} - \dfrac{\beta}{m(i, j)_t}\right)\right)$
    (3.8) $H_{t+1} = H_t + \mu_t H_t \left(x_i - x_j\right)\left(x_i - x_j\right)^T H_t$
Return: H

ALGORITHM 1: Learning Mahalanobis matrix by iterative algorithm.

the aforementioned studies, in this paper, we proposed a fault detection scheme based on metric learning which has been used extensively in the pattern classification problem. The purpose of metric learning is to learn a Mahalanobis distance [17] which can represent an accurate relationship between feature vector and categories of instances. The model focuses on the divergence among classes, instead of extracting the principal components. Meanwhile, the Mahalanobis distance learned from the historical data can be utilized in online detection without real-time update. So, metric learning is more suitable than PCA and FDA for fault diagnosis theoretically.

In practice, selecting an appropriate metric plays a critical role in recent machine learning algorithms. Because the scale of the Mahalanobis distance has no effect on the performance of classification, Mahalanobis distance is the most popular one among numerous metrics. Besides, Mahalanobis distance takes into account of the correlations of different features which can build an accurate distance model. A good metric learning algorithm should be fast and scalable. At the same time, a good metric learning algorithm should emphasize the relevant dimensions while reducing the influence of noninformative dimensions [18]. In this paper, we adopt information-theoretic metric learning (ITML) algorithm to learn Mahalanobis distance function [19]. In ITML algorithm, the distances between similar pairs are bounded in a small given value, while the distances between dissimilar pairs are required to be larger than a large given value in the algorithm. The algorithm is expressed as a particular Bregman optimization problem. To avoid overfitting problem, a method based on LogDet divergence to regularize the target matrix to a given Mahalanobis matrix
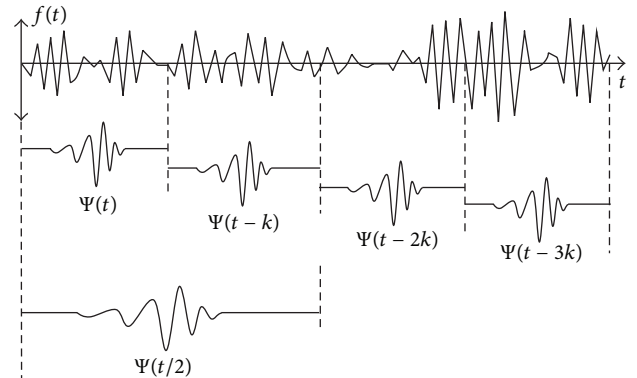

FIGURE 1: Dilating and shifting of wavelet function.

is adopted. It is necessary to remark that a feature extraction method based on wavelet transform is proposed to do the data preprocessing of the algorithm.

The remainder of this paper is organized as follows. In Section 2, we give background knowledge of ITML. Then, wavelet transform is described in Section 3. Section 4 illustrates TE process [20] and gives the experimental results on TE process dataset to demonstrate the good effect of the proposed algorithm. Finally, we draw conclusions and point out future directions in Section 5.

2. Related Work

ITML is a metric learning algorithm without eigenvalue computations or semidefinite programming. And the strategy of
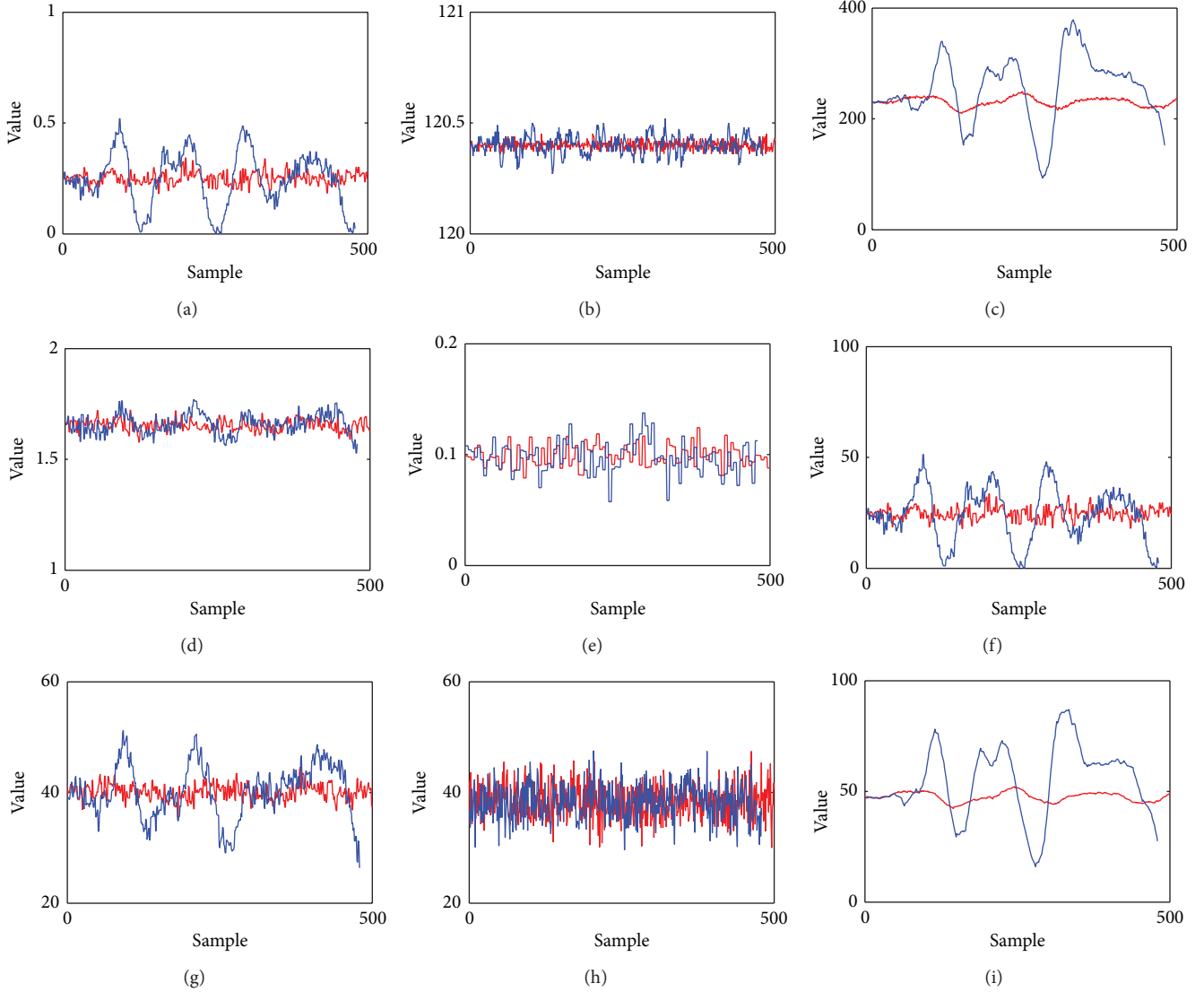
FIGURE 2: Observations of fault 12 in the TE process.

regularizing metric in ITML is to minimize the divergence between the target matrix and a given matrix.

Given a dataset $\{x_k\}$ with $x_k \in R^D$, $k = 1, 2, \ldots, n$. The Mahalanobis distance between x_i and x_j can parameterized by a matrix H as follows:

$$D_H\left(x_i, x_j\right) = \left(x_i - x_j\right)^T H \left(x_i - x_j\right). \tag{1}$$

In ITML, pair constraints are used to represent the relationship of data in the same or different categories. If x_i and x_j are in the same categories, the Mahalanobis distance between them should be smaller than a given value s. Similarly, if x_i and x_j are in different categories, the Mahalanobis distance between them should be larger than a given value b.

The purpose of the ITML is to find a matrix $H \in R^{D \times D}$ which satisfies the following pair constraint sets:

$$
\begin{aligned}
& H \in R^{D \times D}, \\
& \text{s.t. } D_H\left(x_i, x_j\right) \le s\left(i, j\right) \in S, \\
& D_H\left(x_i, x_j\right) \ge b\left(i, j\right) \in D,
\end{aligned} \tag{2}
$$

where S and D represent the set of pairs of data in the same and different categories, respectively.

It deserves pointing out that there will be not only one matrix $H \in R^{D \times D}$ which satisfies all the constraints. To ensure the stability of the metric learning, the target matrix H is regularized to a given function H_0. The distance between H
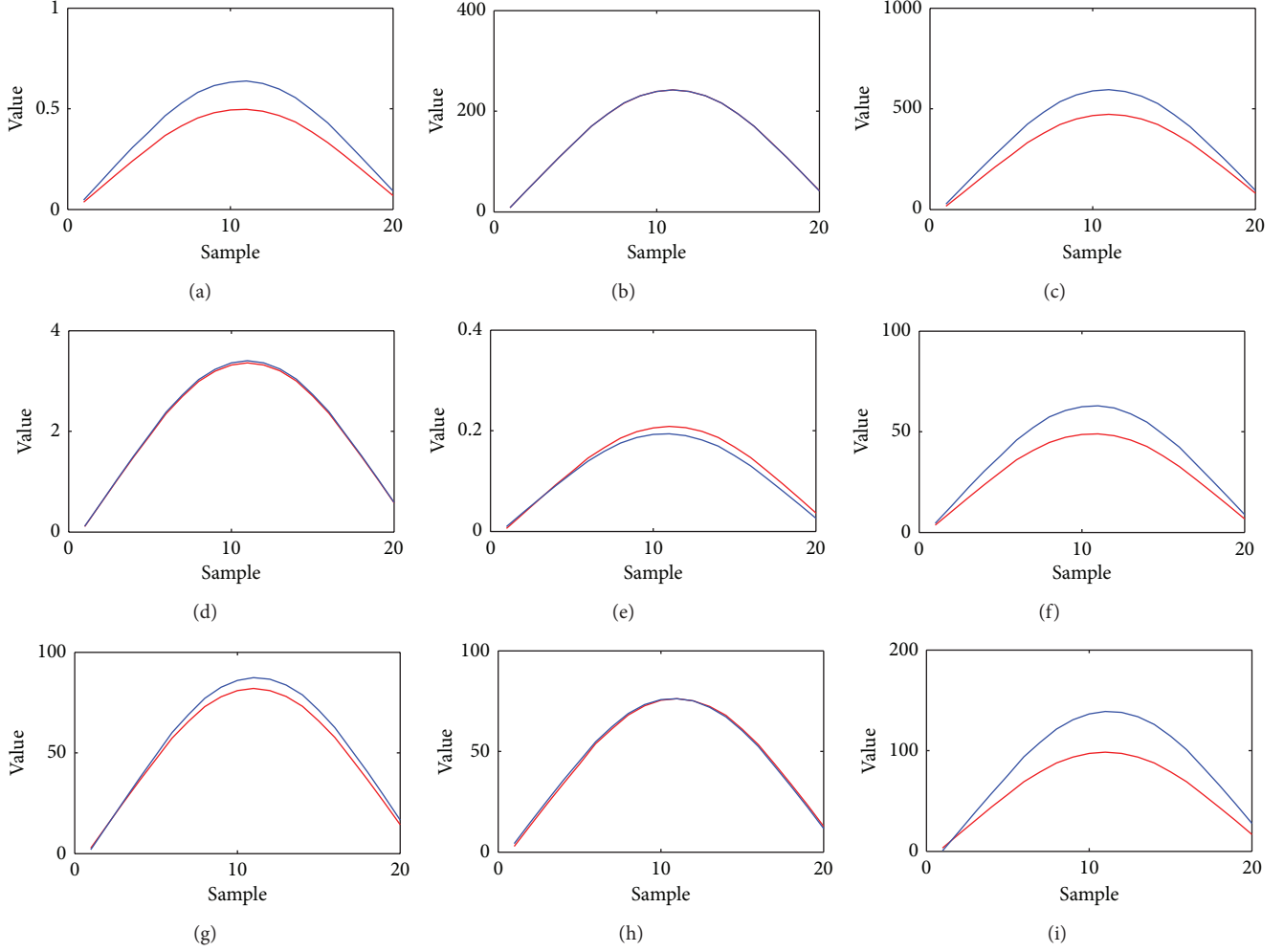
FIGURE 3: Results of the wavelet transform.

and H_0 can be expressed as a type of Bregman matrix divergence [21] as follows:

$$D_\phi (H, H_0) = \phi (H) - \phi (H_0) - \text{tr} \left(\left(\nabla \phi (H_0) \right)^T (H - H_0) \right), \tag{3}$$

in which $\text{tr}(H)$ denotes the trace of matrix H and $\phi(H)$ is a given strictly convex differentiable function that plays a determinant role in the properties of the Bregman matrix divergence. Taking the advantages of different differentiable functions into account, $\phi(H)$ is chosen as $log(\det(H))$. And the corresponding, Bregman matrix divergence $D_\phi(H, H_0)$ is called LogDet divergence. According to the further generalization, the LogDet divergence keeps invariant when performing the invertible linear transformation K, expressed as [22]

$$D_{\text{LD}} (H, H_0) = \text{tr} \left(HH_0^{-1} \right) - \log \left(\det \left(HH_0^{-1} \right) \right) - n,$$

$$D_{\text{LD}} (H, H_0) = D_{\text{LD}} \left(K^T HK, K^T H_0 K \right). \tag{4}$$

The metric learning problem can be translated into a LogDet optimization problem as follows:

$$\min_{H \geq 0} \quad D_{\text{LD}} (H, H_0),$$

$$\text{s.t.} \quad \text{tr} \left(H \left(x_i - x_j \right) \left(x_i - x_j \right)^T \right) \leq s \, (i, j) \in S, \tag{5}$$

$$\text{tr} \left(H \left(x_i - x_j \right) \left(x_i - x_j \right)^T \right) \geq b \, (i, j) \in D.$$

It is worth pointing out that distance constraints $D_H(x_i, x_j) \leq s$ are equivalent to the linear constraints $\text{tr}(H(x_i - x_j)(x_i - x_j)^T) \leq s$. To guarantee the existence of the feasible solution to (5), Kulis proposed an iterative algorithm which introduce slack variable in it [21]. In this way, an iterative equation to update the Mahalanobis distance function is found as follows:

$$H_{t+1} = H_t + \mu H_t \left(x_i - x_j \right) \left(x_i - x_j \right)^T H_t, \tag{6}$$

where μ is a parameter mentioned in Algorithm 1. In the algorithm, the slack variable β balanced the satisfaction of $\min_{H \geq 0} D_{\text{LD}}(H, H_0)$ and the linear constraints. Learning the
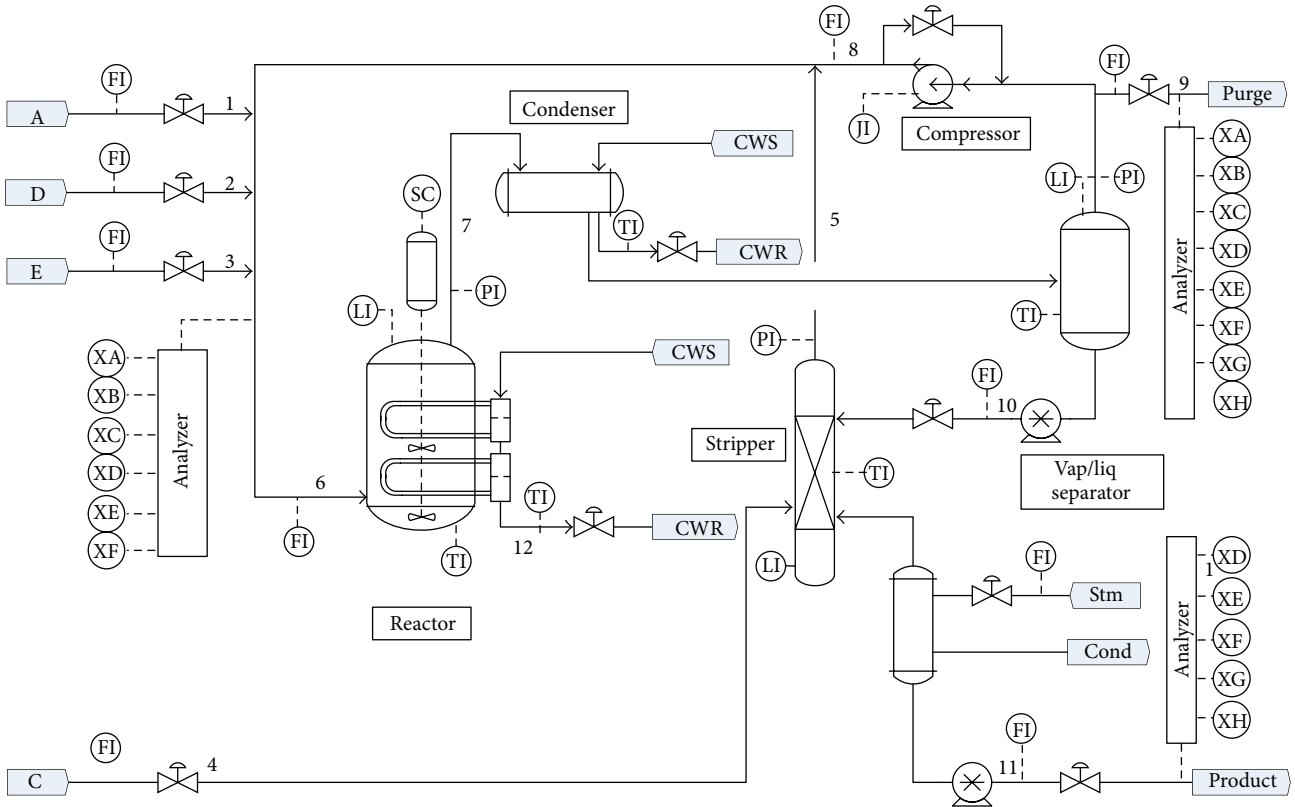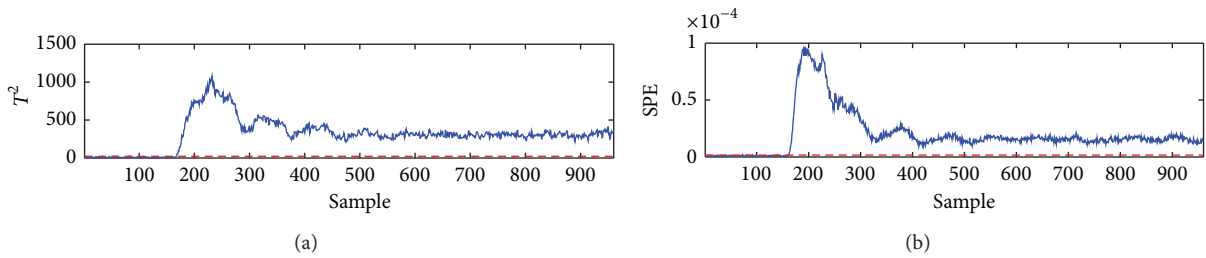
FIGURE 4: The Tennessee Eastman chemical process.

(a)

(b)

FIGURE 5: The classification accuracy of fault 1 dataset for PCA method.

Mahalanobis matrix H based on the given matrix H_0, we can classify the data using k-nearest neighbor classifier to realize failure diagnosis.

3. Fault Diagnosis Using ITML

In the data-driven fault diagnosis system based on the ITML, the system is sensitive to values of the datasets. However, the faults are reflected in vibration amplitude or variation tendency in certain situations. Wavelet transform performs multiscale analysis to the dataset by dilating and shifting the wavelet functions. It transforms the discrepancies of vibration amplitude or variation tendency into the discrepancies of values.

Wavelet functions are localized in time and frequency. Wavelet transform has two main advantages. Firstly, the

analysis window changes itself rather than other complex exponential. Secondly, the duration of the analysis window is not fixed. The wavelet functions are created from the wavelet mother function, by dilating and shifting the window. The wavelet mother function $\psi(t)$ is a function with zero mean which has limited duration and salutatory duration and amplitude. The wavelet functions can be express as [23]

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right), \tag{7}$$

where a is scaling factor and b is translation factor, with $a, b \in R, a \geq 0$. Through increasing the scaling factor a, the wavelet function is expanded and is conducive to analysis signals with low frequency and long duration. Correspondingly, by reducing the scaling factor a, the wavelet function is shrunk and is conducive to analysis signals with high frequency
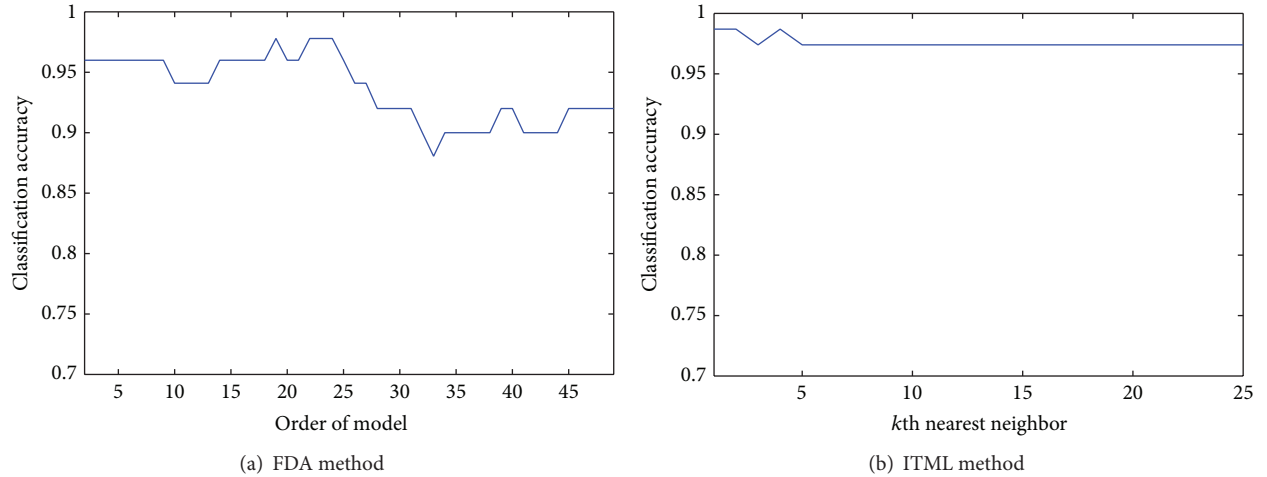
(a) FDA method

(b) ITML method

FIGURE 6: The classification accuracy of fault 1 dataset for 2 methods. (a) The experimental results based on FDA method. (b) The experimental results based on ITML method.

and short duration. By changing the translation factor b, the wavelet functions can realize the traversal along the time axis to get the information of time domain. The wavelet transform can study different scale features and information of time domain which can be expressed as in Figure 1.

The wavelet transform aims at getting a linear combination of the wavelet functions to describe the features in the signal. The value of the wavelet transform is generated by different scaling factors and translation factors. The wavelet transform is defined as [23]

$$ \mathrm{WT}_{a,b} = \frac{1}{\sqrt{a}} \int_R x(t) \psi^* \left(\frac{t-b}{a} \right) dt. \tag{8} $$

Wavelet transform performs multiscale analysis to the dataset which is conducive to the results of ITML. In order to verify this, a wavelet transform to the dataset of TE process is constructed. TE process is introduced in Section 4. Selecting the corresponding 20 consecutive observations of the 9 variables of fault 12 dataset in the TE process randomly, the results of the wavelet transform are shown in Figures 3 and 4. The red lines in Figures 2 and 3 represent the value of fault-free dataset and the blue lines represent the value of fault 12 dataset.

The results of wavelet transform show that features in the signal are converted into the discrepancies of values. Wavelet transform performs well in doing the feature extraction of the ITML.

4. Experimental Results

4.1. Dataset. The designed method of the data-driven fault diagnosis system proposed in this work is applied on the Tennessee Eastman chemical process.

TE process is a chemical plant using as an industrial benchmark process; the schematic flow diagram and instrumentation of which are shown in Figure 4 [24]. TE process gets two products from four reactants. All the 52 variables

TABLE 1: Manipulating variables in TE plant.

Variable number	Variable name
XMV (1)	D feed flow (stream 2)
XMV (2)	E feed flow (stream 3)
XMV (3)	A feed flow (stream 1)
XMV (4)	A and C feed flows (stream 4)
XMV (5)	Compressor recycle valve
XMV (6)	Purge valve (stream 9)
XMV (7)	Separator pot liquid flow (stream 10)
XMV (8)	Stripper liquid product flow (stream 11)
XMV (9)	Stripper steam valve
XMV (10)	Reactor cooling water flow
XMV (11)	Condenser cooling water flow

contained in the process are 11 control variables and 41 measurement variables, respectively, as listed in Table 1 [16] and Table 2 [16].

20 process faults and a valve fault are defined in TE process, as shown in Table 3 [16]. In the work of Chiang et al. [15], a widely used dataset of TE process is given. To copy the measurements of 52 variables for 24 hours, 22 training datasets are contained in the dataset corresponding to the fault-free operating condition and 21 fault operating conditions. Simultaneously, 22 test datasets are contained in the dataset, in which the measurements of 52 variables for 48 hours are collected. It is worth pointing out that the faults in the 22 test datasets are added after 8 simulation hours. The sampling time of both of 22 training datasets and 22 test datasets is 3 minutes.

4.2. Performance Comparing with Classical Methods. To demonstrate the advantages of the proposed fault detection method, we compare it to two classical1 methods, PCA and FDA. We carried out experiments on the dataset of TE

TABLE 2: Measured variables in TE plant.

Variable number	Variable name
XMEAS (1)	A feed (stream 1)
XMEAS (2)	D feed (stream 2)
XMEAS (3)	E feed (stream 3)
XMEAS (4)	A and C feed (stream 4)
XMEAS (5)	Recycle flow (stream 8)
XMEAS (6)	Reactor feed rate (stream 6)
XMEAS (7)	Reactor pressure
XMEAS (8)	Reactor level
XMEAS (9)	Reactor temperature
XMEAS (10)	Purge rate (stream 9)
XMEAS (11)	Product separator temperature
XMEAS (12)	Product separator level
XMEAS (13)	Product separator pressure
XMEAS (14)	Product separator underflow (stream 10)
XMEAS (15)	Stripper level
XMEAS (16)	Stripper pressure
XMEAS (17)	Stripper underflow (stream 11)
XMEAS (18)	Stripper temperature
XMEAS (19)	Stripper steam flow
XMEAS (20)	Compressor work
XMEAS (21)	Reactor cooling water outlet temperature
XMEAS (22)	Separator cooling water outlet temperature
XMEAS (23)	Component A (stream 6)
XMEAS (24)	Component B (stream 6)
XMEAS (25)	Component C (stream 6)
XMEAS (26)	Component D (stream 6)
XMEAS (27)	Component E (stream 6)
XMEAS (28)	Component F (stream 6)
XMEAS (29)	Component A (stream 9)
XMEAS (30)	Component B (stream 9)
XMEAS (31)	Component C (stream 9)
XMEAS (32)	Component D (stream 9)
XMEAS (33)	Component E (stream 9)
XMEAS (34)	Component F (stream 9)
XMEAS (35)	Component G (stream 9)
XMEAS (36)	Component H (stream 9)
XMEAS (37)	Component D (stream 11)
XMEAS (38)	Component E (stream 11)
XMEAS (39)	Component F (stream 11)
XMEAS (40)	Component G (stream 11)
XMEAS (41)	Component H (stream 11)

TABLE 3: Process faults in TE plant.

Fault number	Process variable
IDV (1)	A/C feed ratio, B composition constant
IDV (2)	B composition, A/C ration constant
IDV (3)	D feed temperature
IDV (4)	Reactor cooling water inlet temperature
IDV (5)	Condenser cooling water inlet
IDV (6)	A feed loss
IDV (7)	C header pressure loss-reduced availability
IDV (8)	A, B, C feed composition
IDV (9)	D feed temperature
IDV (10)	C feed temperature
IDV (11)	Reactor cooling water inlet temperature
IDV (12)	Condenser cooling water inlet temperature
IDV (13)	Reaction kinetics
IDV (14)	Reactor cooling water valve
IDV (15)	Condenser cooling water valve
IDV (16)	Unknown
IDV (17)	Unknown
IDV (18)	Unknown
IDV (19)	Unknown
IDV (20)	Unknown
IDV (21)	The valve fixed at steady state position

process and the classification accuracy of k-nearest neighbor is chosen to evaluate the performance of classification.

The experiments are conducted on 6 datasets in the TE process, fault-free dataset, fault 1 dataset, fault 2 dataset, fault 4 dataset, fault 6 dataset, and fault 7 dataset, respectively. The feature extraction method of the datasets of TE process is selected as wavelet transform. To balance the performance of the feature extraction with the amount of delay, every 7 consecutive samples are collected to do a wavelet transform.

The slack variable used to avoid the overfitting problem is set as $\beta = 10^{-3}$ and all results presented are the average over 10 runs. The experimental results of fault 1 dataset are given in Figures 5 and 6.

Figure 5 shows the result of fault detection of fault 1 dataset for PCA method when fault occurs in both of the two orthogonal subspaces, which can be successfully detected by SPE and T^2 statistics. And the fault detection accuracy of fault 1 dataset for PCA method is 0.99. PCA method provides a satisfactory fault detection rate, but it cannot estimate fault types because it determines the lower dimensional subspaces without considering the information between the classes. Figure 6(a) indicates that the classification accuracy of FDA method float in line with the order of model and the classification accuracy are not totally satisfactory. Figure 6(b) illustrates that the ITML method gives higher fault detection rate than FDA method and it remains stable for different kth nearest neighbor. Furthermore, ITML method takes advantages of PCA method that it can estimate fault types directly.

Experimental results are summarized in Figure 7 and these results reveal that ITML method is more robust than PCA and FDA. Considering the ability of estimating fault types directly, ITML method achieves the best classification accuracy across all datasets. And the performance and effectiveness of the wavelet transform based feature extraction are demonstrated by the results of the experiment.
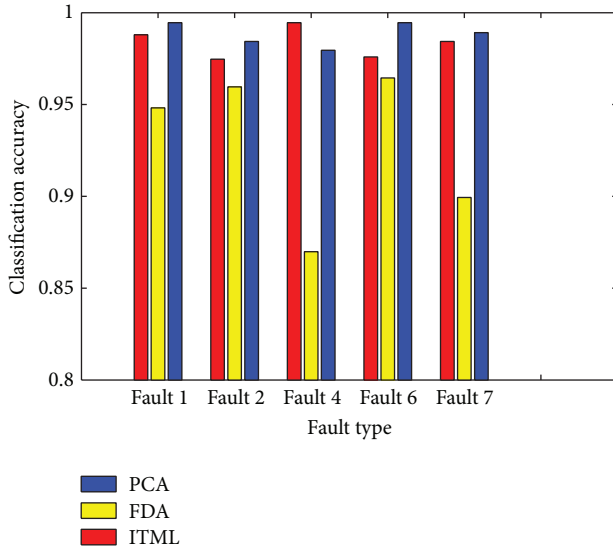
FIGURE 7: The classification accuracy of 5 different datasets for 3 methods.

5. Conclusion

In this paper, we proposed a fault detection scheme based on information-theoretic metric learning. ITML performs well in learning Mahalanobis distance function. In the proposed framework, the feature vector is firstly extracted by applying wavelet transform. After that, we apply the ITML algorithm in fault detection method to improve fault detection accuracy and estimate fault types. Comparing with the fault detection schemes based on PCA and FDA, experiments on TE process dataset demonstrate that the proposed method is more robust. The performance and effectiveness of the wavelet transform-based feature extraction are demonstrated by the results of the experiments at the same time.

Conflict of Interests

The authors declared that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] Z. Xudong, Z. Lixian, S. Peng, and L. Ming, "Stability and stabilization of switched linear systems with mode-dependent average dwell time," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1809–1815, 2012.

[2] Z. Xudong, Z. Lixian, S. Peng, and L. Ming, "Stability of switched positive linear systems with average dwell time switching," *Automatica*, vol. 48, no. 6, pp. 1132–1137, 2012.

[3] Z. Xudong and L. Xingwen, "Improved results on stability of continuous-time switched positive linear systems," *Automatica*, 2013.

[4] S. Yin, S. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic datadriven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.

[5] S. Yin, X. Yang, and H. R. Karimi, "Data-driven adaptive observer for fault diagnosis," *Mathematical Problems in Engineering*, vol. 2012, Article ID 832836, 21 pages, 2012.

[6] S. Yin, H. Luo, and S. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 2402–2411, 2014.

[7] R. Dunia, S. J. Qin, T. F. Edgar, and T. J. McAvoy, "Use of principal component analysis for sensor fault identification," *Computers and Chemical Engineering*, vol. 20, pp. 713–718, 1996.

[8] S. Yin, S. X. Ding, A. H. A. Sari, and H. Hao, "Data-driven monitoring for stochastic systems and its application on batch process," *International Journal of Systems Science*, vol. 44, no. 7, pp. 1366–1376, 2013.

[9] S. Yin, G. Wang, and H. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, 2013.

[10] I. T. Jolliffe, *Principal Component Analysis*, Springer, Berlin, Germany, 1986.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, NY, USA, 2001.

[12] D. Zumoffen and M. Basualdo, "From large chemical plant data to fault diagnosis integrated to decentralized fault-tolerant control: pulp mill process application," *Industrial and Engineering Chemistry Research*, vol. 47, no. 4, pp. 1201–1220, 2007.

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

[14] N. D. Tracy, J. C. Young, and R. L. Mason, "Multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 24, no. 2, pp. 88–95, 1992.

[15] 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," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 2, pp. 243–252, 2000.

[16] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, Springer, London, UK, 2001.

[17] S. Xiang, F. Nie, and C. Zhang, "Learning a mahalanobis distance metric for data clustering and classification," *Pattern Recognition*, vol. 41, no. 12, pp. 3600–3612, 2008.

[18] L. Meizhu and B. Vemuri, "A robust and efficient doubly regularized metric learning approach," in *Computer Vision—ECCV 2012*, pp. 646–659, Springer, Berlin, Germany, 2012.

[19] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proceedings of the 24th International Conference on Machine learning (ICML '07)*, pp. 209–216, ACM, June 2007.

[20] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers and Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993.

[21] B. Kulis, M. Sustik, and I. Dhillon, "Learning low-rank kernel matrices," in *Proceedings of the 23th International Conference on Machine Learning (ICML '06)*, pp. 505–512, ACM, June 2006.

[22] J. V. Davis and I. Dhillon, "Differential entropic clustering of multivariate gaussians," *Advances in Neural Information Processing Systems*, vol. 19, p. 337, 2007.

[23] T. A. Ridsdill-Smith, *The Application of the Wavelet Transform to the Processing of Aeromagnetic Data*, The University of Western Australia, Crawley, Australia, 2000.

[24] S. Yin, *Data-Driven Design of Fault Diagnosis Systems*, VDI, Dusseldorf, Germany, 2012.