Journal of Manufacturing Systems 61 (2021) 536–545

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Quality-relevant fault detection based on adversarial learning and distinguished contribution of latent variables to quality

Shifu Yan, Xuefeng Yan *

Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China

A R T I C L E I N F O

Keywords:
Quality relevant
Fault detection
Nonlinear process
Adversarial learning
Local projection

A B S T R A C T

Quality-relevant fault detection is a primary task to reveal the changes of quality variables in process monitoring. Current works mainly focus on learning quality-relevant features, however, how to distinguish quality-relevant and irrelevant information is responsible for the excellent monitoring performance. In this study, a novel quality-relevant fault detection method is proposed on the basis of adversarial learning and distinguished contribution of latent features to quality is originally introduced. First of all, we map the input variables into a gaussian manifold in adversarial and unsupervised manner. Then a fully connected neural network is trained to learn the relationship between latent and quality variables. To distinguish necessary information in such manifold, the Jacobi operator at the corresponding point is calculated to project the latent variables into quality-relevant and quality-irrelevant subspaces. Third, fault detection is implemented in these dynamic subspaces using the probabilities of latent variables. Finally, the proposed method is evaluated by numerical example, the Tennessee-Eastman process and wind turbine blade icing process.

1. Introduction

With the expansion of the modern industrial scale, a real-time monitoring method is required to ensure the safety and economics of industrial processes. Currently, data-driven approaches have been extensively studied in industrial processes instead of traditional model-based methods [1–5]. In all the monitoring tasks, quality-relevant fault detection has been a popular research topic in recent years; this technique aims to reveal the abnormal status of quality variables that are directly related to the benefits of industries [6]. Therefore, monitoring such variables can effectively reduce the unnecessary downtime of devices and improve the economics of enterprises based on production safety.

Quality variables are typically difficult, costly to measure, and have a large time delay. Values of these variables are commonly predicted using related process variables based on models like least squares (LS) and partial least squares (PLS). For improved performance, the traditional PLS-based method was successively extended into total PLS, which includes four subspaces by Zhou et al. [7] and concurrent PLS, which includes three subspaces by Qin et al. [8]. To avoid collinearity, principal component analysis (PCA) or kernel PCA can be used to extract quality-relevant latent variables as well as orthogonal signal correction and dynamic ideas [9–12]. However, these methods mainly rely on the linear hypothesis, which is inefficient in current industries. Thus, kernel methods are widely used for nonlinear processes such as total kernel PLS (TKPLS) [13,14]. Although the kernel-based methods learn the nonlinear representations of the data, these methods mainly focus on the nonlinear relationship within the process variables and neglect the nonlinear relationship between the process and quality variables and different variables under normal or faulty conditions contribute to quality variables differently in nonlinear cases. Therefore, how to extract the most quality-relevant information is the key for accurate detection performance.

Alternatively, neural networks (NNs) are among the efficient tools for nonlinear tasks on the basis of their outstanding performance in representation learning [15]. For process monitoring, stack autoencoders and deep belief nets (DBNs) are two extensively used models and the status of industrial processes is usually monitored by Mahalanobis or Euclid distance [16,17]. Data in current industries are typically complex, and these methods ignore the real distribution of process variables that can be learned using generative models. However, these models such as Gaussian mixture models mainly based on Markov chain Monte

* Corresponding author at: P.O. BOX 293, East China University of Science and Technology, MeiLong Road 130, Shanghai 200237, PR China.
E-mail address: xfyan@ecust.edu.cn (X. Yan).

https://doi.org/10.1016/j.jmsy.2021.10.001
Received 21 June 2021; Received in revised form 26 September 2021; Accepted 3 October 2021
Available online 19 October 2021
0278-6125/

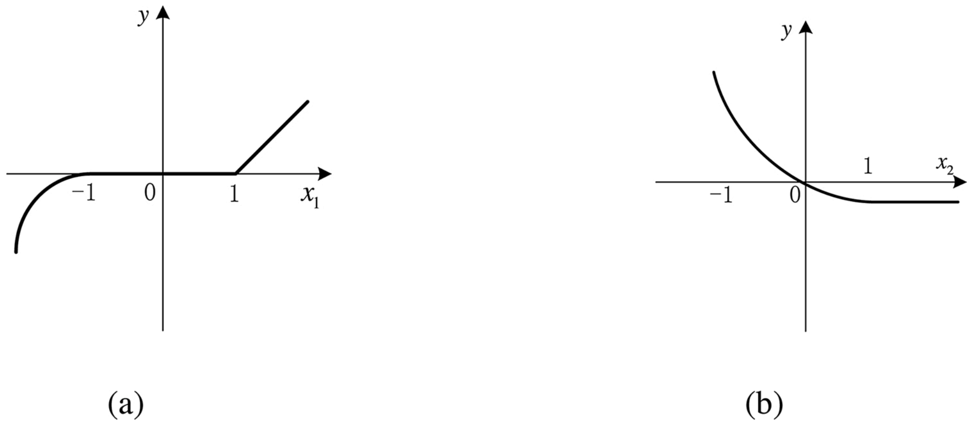

Fig. 1. Simple example for revealing the motivation. (a) Relationship between y and x_1 when $x_2 = 0$; (b) relationship between y and x_2 when $x_1 = 0$.

Carlo (MCMC) methods that suffer from costly computation [18,19]. To avoid such complicated approximation or inference, Generative adversarial nets (GANs), which model distributions by adversarial learning between two NNs is proposed [20]. In an unsupervised way, adversarial autoencoder (AAE) is presented to learn the mapping function between the original data to a specific manifold [21]. The GAN-based method has been extensively used in anomaly detection given its capability for learning the original distribution [22,23].

Based on the abovementioned description, a novel method for nonlinear process based on adversarial learning and local projection (ALLP) is proposed in this study. Firstly, input variables are scaled to a predefined manifold by adversarial learning. Second, the correlation between these latent variables and corresponding quality variables are modeled by a fully connected NN (FCNN). Then, local projection is presented to distinguish different contributions of latent variables to quality. This process can be implemented by calculating the Jacobi operator combined with singular value decomposition (SVD), and the quality-relevant and quality-irrelevant subspaces are further established under different conditions of latent variables. Thus, the probabilities of these latent variables can be estimated in these two dynamic subspaces to monitor the processes.

The contributions of this work can be summarized as:

1) the different contribution of latent variables to predict quality is firstly studied;

2) a dynamic subspace decomposition method is proposed to extract the most influential information of process variables based on the Taylor expansion;

3) a direct scheme to detect quality-relevant faults in dynamic subspaces to conquer the unknowingness of data distribution.

Notation: Let $\Re^{n\times m}$ be the dataset with n samples of m-dimensional variables and $\Re^m$ be the set of m-dimensional vector. $N(\mu,\delta^2)$ is the nomenclature between 0 and 1. $(\cdot)_{new}$ indicates new sample. J_{th} is the predefined threshold of T^2 statistic which is calculated by mahalanobis distance. $(\cdot)_y$ and $(\cdot)_o$ is this manuscript represent quality-relevant and quality-irrelevant symbols. ∇ is the gradient of network training. $span\{\cdot\}$ and $span\{\cdot\}^{\perp}$ define two spaces that are orthogonal to each other.

2. Background

2.1. Motivation

For quality-relevant fault detection, the common procedure is to extract the quality-relevant features from the process variables. Traditional linear methods consider the correlation to be fixed in obtaining quality-relevant information, as reflected in linear projection vectors [7–11]. In current industries, a nonlinear relationship inevitably exists in process and quality variables. Traditional nonlinear methods only consider the nonlinearities within process variables and also extract the quality-relevant information in a constant way [13,14]. However, the contribution of process variables will vary under different conditions, which can be illustrated by a simple example as follows.

Suppose a quality variable y to be related to two variables x_1, x_2, i.e. $y = f(x_1, x_2)$. For ease of description, the marginal relationship between y and x_1 or x_2 is depicted in Figs. 1(a) and (b) where x_1 and x_2 do not contribute to the variety of y in the full space of variables. For example, y is both related to x_1 and x_2 in the neighborhood of $(x_1 = -1.5, x_2 = -1.5)$. However, near the point $(x_1 = 1.5, x_2 = 1.5)$ or $(x_1 = 0, x_2 = 0)$, y is related to only x_1 or x_2.

Therefore, it is necessary to extract the most quality-related information of process variables during the faulty situations to reach accurate monitoring performance. In nonlinear processes, this type of contribution of process variables to quality is defined as distribution contribution in the following description. Given this idea, we propose the following local projection method to obtain the distributed contribution under different conditions and dynamically extract quality-related information.

2.2. Background of quality relevant fault detection

Process variables under normal conditions are collected as $\mathbf{x} \in \Re^m$, and a new variable $\mathbf{x}_{new}$ that has been affected can be described as

$$\mathbf{x}_{new} = \mathbf{x} + \mathbf{\Theta} j, \tag{1}$$

where $\mathbf{\Theta} \in \Re^m$ and j correspond to the faulty direction and magnitude. To detect this disturbance, hypothesis testing is used where null hypothesis indicates fault-free, $f = 0$; and alternative hypothesis indicates faulty, $f \neq 0$.

Statistics, such as Hotelling's T^2, is constructed as

$$T^2 = \mathbf{x}_{new}^T \mathbf{\Sigma}_{\mathbf{x}}^{-1} \mathbf{x}_{new}, \tag{2}$$

where $\mathbf{\Sigma}_{\mathbf{x}}$ is the covariance of samples under normal condition.

Given a significant level α, the threshold of T^2 can be determined as J_{th} based on $\chi^2_\alpha(d)$, where $d = rank(\mathbf{\Sigma}_{\mathbf{x}})$. If $T^2 \leq J_{th}$, then $f = 0$ is accepted; otherwise, $f \neq 0$ is considered.

To evaluate the monitoring performance, false alarm rate (FAR) and fault detection rate (FDR) are used and defined as

$$\begin{cases} FAR = prob\{T^2 > J_{th} | f = 0\} \\ FDR = prob\{T^2 > J_{th} | f \neq 0\} \end{cases} \tag{3}$$

From the disturbance detected in the process variables, quality-relevant fault denotes the effects that involve quality variable y. Therefore, the relationship between process and quality variables must be established to predict quality variables. In this way, statistics of

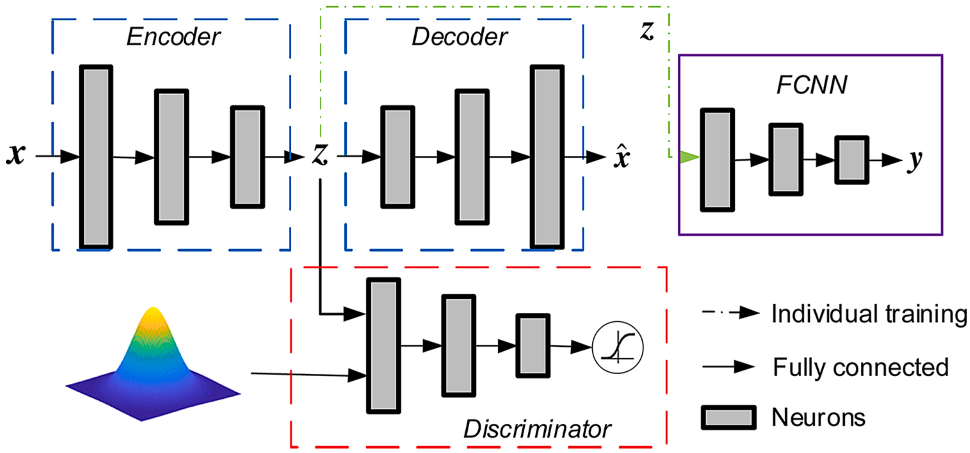

Fig. 2. Model used in the proposed method. The AAE consists of an encoder, a decoder, and a discriminator, which extracts Gaussian variables **z**. The latent **z** is modeled by an FCNN to predict the quality variable *y*. The networks in this study are all fully connected.

process variables are further divided into quality-relevant and quality-irrelevant statistics (T_y^2 and T_o^2) together with corresponding thresholds ($J_{th,y}$ and $J_{th,o}$). The diagnosed logic is described as

$$\begin{cases} T_y^2 \leq J_{thy} \text{ and} T_o^2 \leq J_{tho} \Rightarrow x_{new} \text{ is fault-free} \\ T_y^2 > J_{thy} \Rightarrow \text{quality relevant fault occur} \\ T_y^2 \leq J_{thy} \text{ and} T_o^2 > J_{tho} \Rightarrow \text{quality irrelevant fault occur} \end{cases} \quad (4)$$

2.3. Adversarial autoencoder

As a new paradigm for estimating generative models, GAN consists of a generator (*G*) and a discriminator (*D*). *G* is trained to map random variables **r** into the original **x**, whereas *D* must distinguish whether the data are sampled from the original data or mapped from random variables. The overall objective can be written as follows:

Based on the traditional GAN, AAE is presented to match the distributions of latent variables to an arbitrary prior distribution using the encoder–decoder structure as the generator. As shown in Algorithm I, the encoder (*En*) and decoder (*De*) are optimized by the reconstructive loss in (6) wheras *En* and *D* are trained by the adversarial loss in (7).

$$\min_{En,De} \psi_{re} = -E_{\mathbf{x}}\{\log[p(De(En(\mathbf{x}))\,|\mathbf{x})]\}, \quad (6)$$

$$\min_{En}\max_{D} \psi_{adv} = E_{\mathbf{x}}\{\log[D(\mathbf{z})]\} + E\{\log[1 - D(En(\mathbf{x}))]\} \quad (7)$$

Algorithm I. Training phase of the AAE. The number of steps to update the abovementioned three models can be set differently until converge.

Begin
For $i = 1$ **to** number of epochs **do**
For $j = 1$ **to** number of batches **do**
1) Sample m observations $[\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_m]^T$ from the original data;
2) Sample m random samples $[r_1, r_2, ..., r_d]^T$ from $N(0,1)$;
3) Update the parameters of the encoder and decoder using
$$\nabla = -\frac{1}{m}\sum_{i=1}^{m}\left[\log\left(p\left(De\left(En(\boldsymbol{x}_i)\right)\right)\middle|\boldsymbol{x}_i\right)\right];$$
4) Update the parameters of the discriminator using
$$\nabla = -\frac{1}{m}\sum_{i=1}^{m}\left[\log\left(D(r_i)\right) + \log\left(1 - D\left(En(\boldsymbol{x}_i)\right)\right)\right];$$
5) Update the encoder using
$$\nabla = -\frac{1}{m}\sum_{i=1}^{m}\left[\log\left(1 - D\left(En(\boldsymbol{x}_i)\right)\right)\right].$$
End for
End for
End

$$\min_{G}\max_{D} \psi = E_{\mathbf{x}}[\log D(\mathbf{x})] + E_{\mathbf{z}_r}[\log(1 - D(G(\mathbf{r})))], \quad (5)$$

where $\mathbf{x} \sim p_{\mathbf{x}}$ indicate the original data, and $\mathbf{r} \sim p_r$ are the random variables. Theoretically, GAN can reach the global optimality called Nash equilibrium, where $G(\mathbf{r}) \sim p_g$.

3. Proposed method

In this section, the implementation of the proposed method,

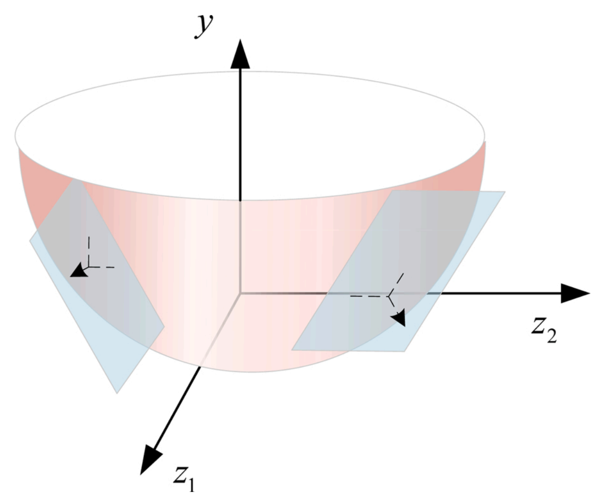

Fig. 3. Illustration of local projection.

including model construction, local projection, and fault detection, is elucidated.

3.1. Model overview

The overall model is depicted in Fig. 2. The AAE is first constructed to map the original variables $\mathbf{x}$ to a Gaussian manifold, in which latent variables are supposed to be $\mathbf{z} \sim N(\mathbf{0}, I_v)$. Thus, the relationship between $\mathbf{x}$ and $\mathbf{z}$ can be obtained using the encoder and decoder of the AAE. In contrast to traditional GAN-based methods, the decoder used in our method is only for improved representation learning of the original variables. On the basis of the extracted latent variables, a nonlinear relationship is built between $\mathbf{z}$ and quality variable y through an FCNN. Sufficient neurons and nonlinear activation functions provide NN with the ability to fit any arbitrarily complex function. The training of the AAE and FCNN is separate; consequently, the feature extraction is unguided by the quality information, which can improve the generalized performance.

3.2. Local projection

Traditional method usually obtains the quality-relevant information of the process variables in a linear way, however, the distribution contribution to quality indicators varies under different conditions in nonlinear processes. The projection vectors that reveal the most influential information to quality variables are dynamic. Thus, the local projection is firstly proposed to extract the most quality-relevant information based on the Taylor expansion at different values, and dynamic decomposition of process space can be realized using Jacobi operator compared with traditional methods. For ease of understanding, the most quality-relevant subspace can be indicated by the tangent plane at the different conditions of variables in Fig. 3.

On the basis of the models constructed in Section III.A, quality variable y can be described as

$$y = f(\mathbf{z}) + \xi, \tag{8}$$

where $f(\cdot)$ is the fitting function constructed using an FCNN, and ξ is the fitting errors of the training samples.

Given a new data point $\mathbf{x}_{new}$, the latent variables are extracted as $\mathbf{z}_{new}$ using the well-trained AAE. The probability of $\mathbf{z}_{new}$ in the latent manifold can be estimated. However, some information on the extracted features contributes to the quality variables. Thus, $\mathbf{z}_{new}$ is projected into quality-relevant and quality-irrelevant subspaces in the following steps.

Since $f(\cdot)$ is constructed by the neural network, it is adequately smooth and differentiable to conduct the backpropagation training.

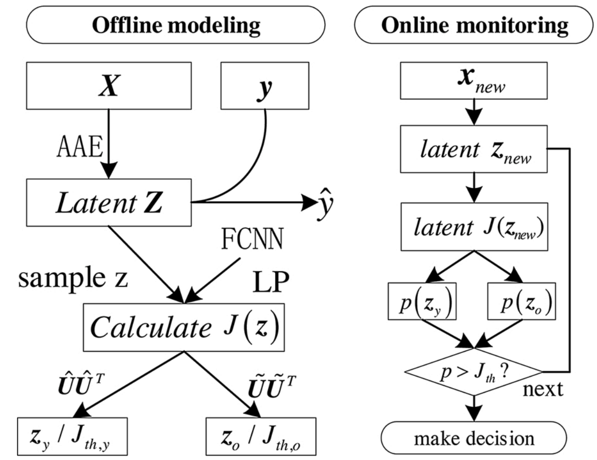

Fig. 4. Flowchart of the proposed methodology.

Thus, the Taylor expansion of this function can be implemented at $\mathbf{z}_{new}$ as

$$f(\mathbf{z}) = f(\mathbf{z}_{new}) + J_f(\mathbf{z}_{new})(\mathbf{z} - \mathbf{z}_{new}) + O\left(\|z - \mathbf{z}_{new}\|^2\right), \tag{9}$$

where $J_f(\mathbf{z}_{new}) \in \Re^{1\times v}$ is the Jacobi operator computed at $\mathbf{z}_{new}$. Thus, the tangent plane that is evidently quality-relevant is spanned by the vectors of $J_f(\mathbf{z}_{new})$. Then, we aim to obtain the quality-relevant contribution of the process variables $\mathbf{z}_{y,new} \equiv span\{J_f(\mathbf{z}_{new})\}$ and quality-irrelevant contribution $\mathbf{z}_{o,new} \equiv span\{J_f(\mathbf{z}_{new})\}^{\perp}$.

Let $\mathbf{M} = J_f(\mathbf{z}_{new})$, and SVD is performed on $\mathbf{M}\mathbf{M}^T$ as

$$\mathbf{M}\mathbf{M}^T = \begin{bmatrix}\widehat{\mathbf{U}} & \widetilde{\mathbf{U}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\widehat{\mathbf{U}}^T\\ \widetilde{\mathbf{U}}^T\end{bmatrix} = \widehat{\mathbf{U}}\Sigma\widehat{\mathbf{U}}^T, \tag{10}$$

where $\widehat{\mathbf{U}} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{pc}]$, $\widetilde{\mathbf{U}} = [\mathbf{u}_{pc+1}, \mathbf{u}_{pc+2}, ..., \mathbf{u}_m]$, and pc is the number of nonzero singular values. The quality-relevant subspace at this data point is spanned by $\widehat{\mathbf{U}}$. The properties of $\widehat{\mathbf{U}}$ and $\widetilde{\mathbf{U}}$ can be described as

$$\widehat{\mathbf{U}}\widehat{\mathbf{U}}^T + \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^T = \mathbf{I}_f, \widehat{\mathbf{U}}^T\widetilde{\mathbf{U}} = 0. \tag{11}$$

Quality-relevant and quality-irrelevant subspaces can be calculated as

$$\mathbf{z}_{y,new} = \mathbf{z}_{new}\widehat{\mathbf{U}}\widehat{\mathbf{U}}^T, \tag{12}$$

$$\mathbf{z}_{o,new} = \mathbf{z}_{new}\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^T. \tag{13}$$

Thus, the two parts are orthogonal to each other because

$$\mathbf{z}_{y,new}\mathbf{z}_{o,new}^T = \mathbf{z}_{new}\widehat{\mathbf{U}}\widehat{\mathbf{U}}^T\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^T\mathbf{z}_{new}^T = 0. \tag{14}$$

The abovementioned local projection method is based on the piecewise linear decomposition. Thus, this method is an extension of the linear version, in which the relationship between z and y is linear. The establishment of the quality-relevant and quality-irrelevant subspaces is dynamic and adaptive for nonlinear processes.

3.3. Scheme of fault detection

The process variable $\boldsymbol{x}$ has been nonlinearly projected into quality-relevant and quality-irrelevant subspaces $\mathbf{z}_y$ and $\mathbf{z}_o$ in a Gaussian manifold. Their probabilities can be estimated as p_y and p_o that are treated as the statistics in comparison with traditional fault detection methods. The training dataset is used to calculate the thresholds $J_{th,y}$ and $J_{th,o}$ of p_y and p_o. Thus, the entire fault detection approach is summarized

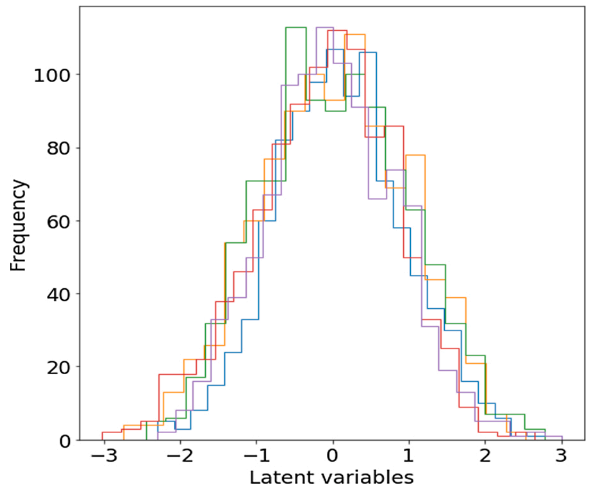

Fig. 5. Frequency distribution histogram of the latent variables extracted using the AAE in the numerical example.

in Algorithm II and the data flow is illustrated in Fig. 4. Based on the features extracted by adversarial learning, the unknowingness of data distribution can be avoid and the fault detection can be implemented in a more direct and interpretable way.

Remark: The following experiments are run in a computer with Intel Xeon Silver 4110 2.1 GHz, 64 GB RAM. The training of AAE and FCNN is executed using a mini-batch stochastic gradient descent optimizer with the learning rate $lr = 0.01$ for the encoder, decoder, and FCNN and $lr = 0.005$ for the discriminator. "tanh," "linear," and "rectified linear unit (ReLU)" are used respectively as the activations of the encoder, decoder, and discriminator. Thresholds for statistics are calculated under a significant level $\alpha = 0.99$. In the proposed scheme, each collected sample can be preprocessed within 2 ms which meets the real-time requirement.

Algorithm II. Implementation of ALLP-based fault detection scheme.

Off-line modeling:

1) Collect and then normalize n samples $\boldsymbol{X} \in \Re^{n\times m}$ and $\boldsymbol{y} \in \Re^{n}$.
2) Construct an AAE model using $\boldsymbol{X}$, and extract the latent variables $\boldsymbol{Z} \in \Re^{n\times v}$.
3) Construct the FCNN between $\boldsymbol{Z}$ and $\boldsymbol{y}$.
4) **For** each sample $\boldsymbol{z}$ in the training dataset **do**
    (4.1) Calculate the Jacobi operator $J_f(\boldsymbol{z})$;
    (4.2) Calculate the projection matrix using (10–14).
    (4.3) Calculate the p_y and p_o of $\boldsymbol{z}_y$ and $\boldsymbol{z}_o$.
   End for
5) Given significant level α, thresholds $J_{th,y}$ and $J_{th,o}$ are calculated through kernel density estimation.

Online monitoring:

1) Collect and then normalize a new sample $\boldsymbol{x}_{new}$.
2) Obtain the latent features $\boldsymbol{z}_{new}$ using the constructed AAE.
3) Calculate $J_f(z_{new})$ and the projection matrix using (10–14).
4) Estimate the $p_{y,new}$ and $p_{o,new}$ of $\boldsymbol{z}_{y,new}$ and $\boldsymbol{z}_{o,new}$.
5) Make decisions in accordance with (5).

Table 1
FDRs (%) statistics of TKPLS, MKPLS and ALLP for numerical example.

Fault No.	TKPLS				MKPLS		ALLP	
	T_y^2	T_o^2	T_r^2	Q	T_y^2	T_o^2	p_y	p_o
1	32.6	100	100	0	32.6	100	100	100
2	0.3	98.6	100	0.4	0.3	100	0.3	100

4. Simulation studies

4.1. Numerical example

A numerical example used to evaluate the method is designed as follows:

$$\begin{cases} x_1 = t_1 + e_1 \\ x_2 = t_1 + t_2^2 + e_2 \\ x_3 = t_3 + t_4 + e_3 \\ x_4 = t_3^2 - 2t_5 + e_4 \\ x_5 = t_5 - t_5^3 + e_5 \\ x_6 = t_1 + \sin(t_2) + e_6 \\ x_7 = t_2 - t_3 + e_7 \\ y = x_3 + x_2x_4 + e_8 \end{cases}, \tag{15}$$

where $t_i \sim U(0,1), i = 1,2,...,5$ are the latent variables that form this example. Gaussian noises $e_i \sim N(0, 0.1^2), i = 1,2,...,8$ have been added during the data generation to simulate more real-world data. Seven variables are included, and the quality is relevant to x_2, x_3, x_4. Thus, three datasets with 1000 samples in each set are generated, and samples without disturbance are used for training. Simultaneously, two types of faults are introduced in the remaining two datasets. Fault 1 is a step deviation of 2.5 in x_2 at the 201 st point. Fault 2 is a step deviation of 2.5 in x_7 at the 201 st point.

Faults 1 and 2 are quality-relevant and quality-irrelevant, correspondingly. An AAE consists of 7−5-7 autoencoder with a nonlinear encoder and a linear decoder, and 7−10-1 discriminator is used to extract the latent variables. The distributions of these latent variables, which are nearly normally distributed, are demonstrated in Fig. 5. These

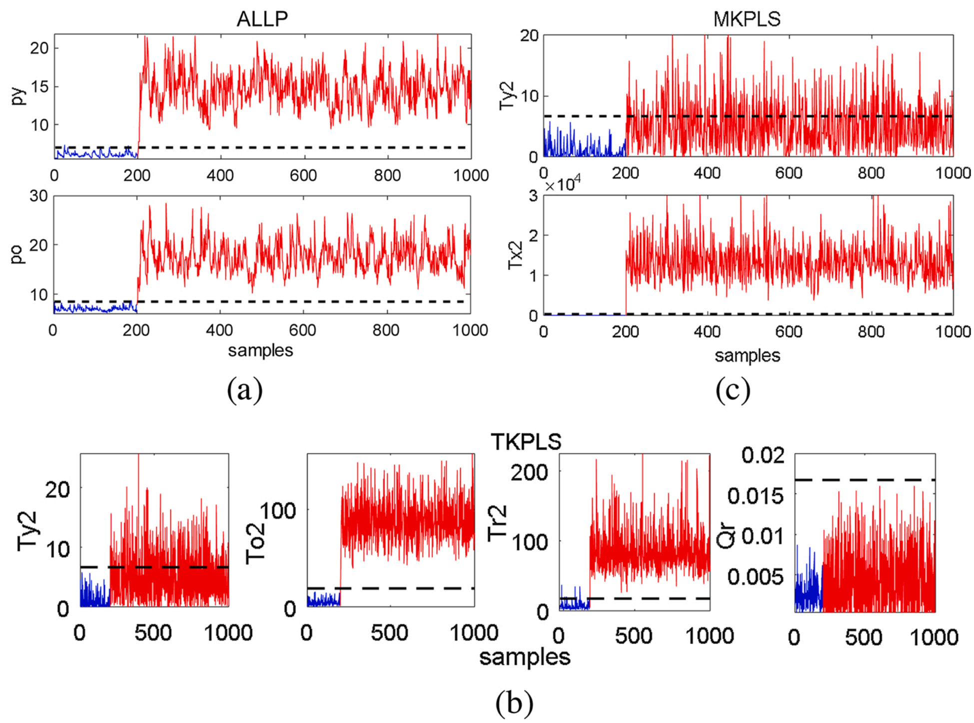

Fig. 6. Monitoring results of fault 1 in numerical example. (a) the results of proposed method including statistics p_y and p_o; (b) the results of TKPLS including statistics T_y^2, T_o^2, T_r^2, Q; (c) the results of MKPLS including statistics T_y^2, T_x^2.

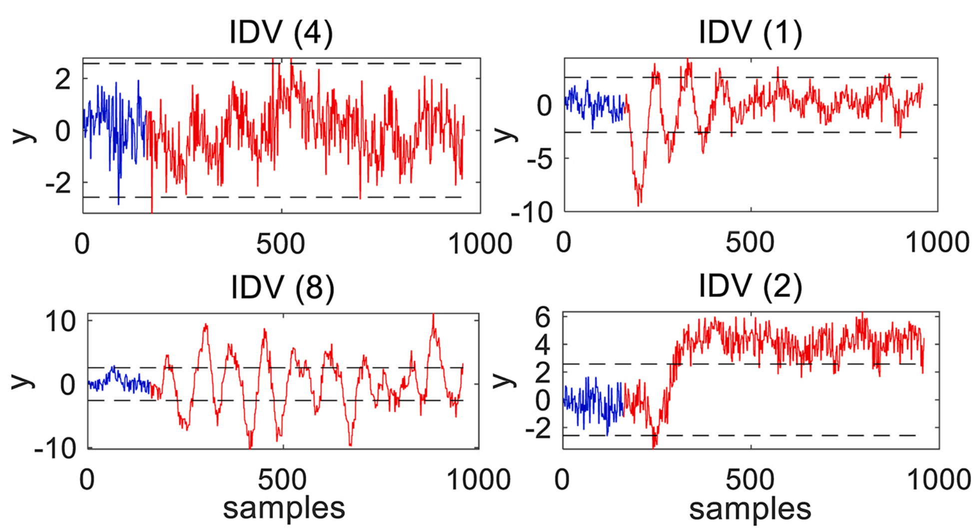

Fig. 7. Real values of y in IDV (1), (2), (4), and (8).

Table 2
FDRs of statistics of KPCR, KPLS, MKPLS, and ALLP for quality relevant faults in TEP.

Fault No.	Real	KPCR		KPLS		TKPLS				MKPLS		ALLP	
	n_t	T_y^2	T_o^2	T_y^2	T_o^2	T_y^2	T_o^2	T_r^2	Q	T_y^2	T_o^2	p_y	p_o
1	14.6	38.1	99.5	99.6	99.8	96.9	99.6	98.8	99.3	96.9	**99.8**	24.4	99.6
5	9.6	30.8	23.9	24.9	21.6	10.8	24.6	9.6	22.0	10.8	**100**	4.5	40.4
7	13.1	40.9	**100**	99.0	**100**	55.8	97.3	91.0	95.9	55.8	**100**	42.5	**100**
2	65.2	87.8	98.3	98.0	98.0	89.5	98.0	98.0	97.0	89.5	**98.5**	93.1	98.3
6	79.6	10.0	99.5	99.5	**100**	97.5	98.8	97.8	**100**	97.5	**100**	85.9	99.6
18	71.5	19.8	89.6	89.0	89.4	87.6	88.6	87.4	89.1	87.6	89.8	86.9	**90.4**
21	20.4	50.0	55.9	51.9	34.0	54.5	37.6	6.1	32.3	54.5	45.4	30.1	**56.1**
8	44.8	89.4	97.6	96.9	97.1	64.9	94.4	81.6	95.0	64.9	97.6	46.6	**98.9**
12	49.8	87.3	**99.8**	98.8	97.9	68.0	97.3	74.4	97.0	68.0	99.4	38.5	**99.8**
13	52.3	84.8	95.0	94.6	93.8	86.8	94.5	70.8	94.3	86.8	95.1	71.5	**95.5**

Table 3
FDRs of statistics of KPCR, KPLS, MKPLS, and ALLP for quality irrelevant faults in TEP.

Fault No.	Real	KPCR		KPLS		TKPLS				MKPLS		ALLP	
	n_t	T_y^2	T_o^2	T_y^2	T_o^2	T_y^2	T_o^2	T_r^2	Q	T_y^2	T_o^2	p_y	p_o
3	2.7	13.8	1.5	2.5	0.3	1.5	1.6	0.1	0	1.5	0.6	1.0	**17.88**
4	1.3	8.5	95.3	66.6	60.9	20.4	48.8	0.3	14.0	20.4	99.8	0	**100**
9	1.5	11.1	3.6	1.1	0.1	2.0	0.9	0.4	0	2.0	0.4	0	**16.88**
10	5.0	36.9	52.1	82.4	35.6	62.6	**79.4**	2.4	32.9	62.6	69.4	0.1	70.63
11	1.3	13.8	86.1	54.8	48.3	19.4	52.3	2.1	36.6	19.4	67.1	0	**97**
14	0.6	2.5	**100**	98.5	99.9	70.5	98.6	33.1	99.9	70.5	**100**	3.3	**100**
15	1.3	15.5	4.4	6.25	0.1	4.9	3.8	0	1.0	4.9	2.5	0	**27.38**
16	2.7	28.6	30.3	63.4	15.0	38.6	51.0	0.3	16.4	38.6	65.3	0.3	**67.75**
17	4.6	11.3	96.8	82.3	84.8	48.8	79.8	37.0	80.5	48.8	94.3	6.1	**97.13**
19	1.3	3.3	17.9	2.8	2.5	1.4	1.6	1.3	0	1.4	30.5	0	**43.75**
20	3.4	35.6	59.9	37.1	38.5	16.1	33.3	2.8	36.0	16.1	57.9	0.8	**89.5**

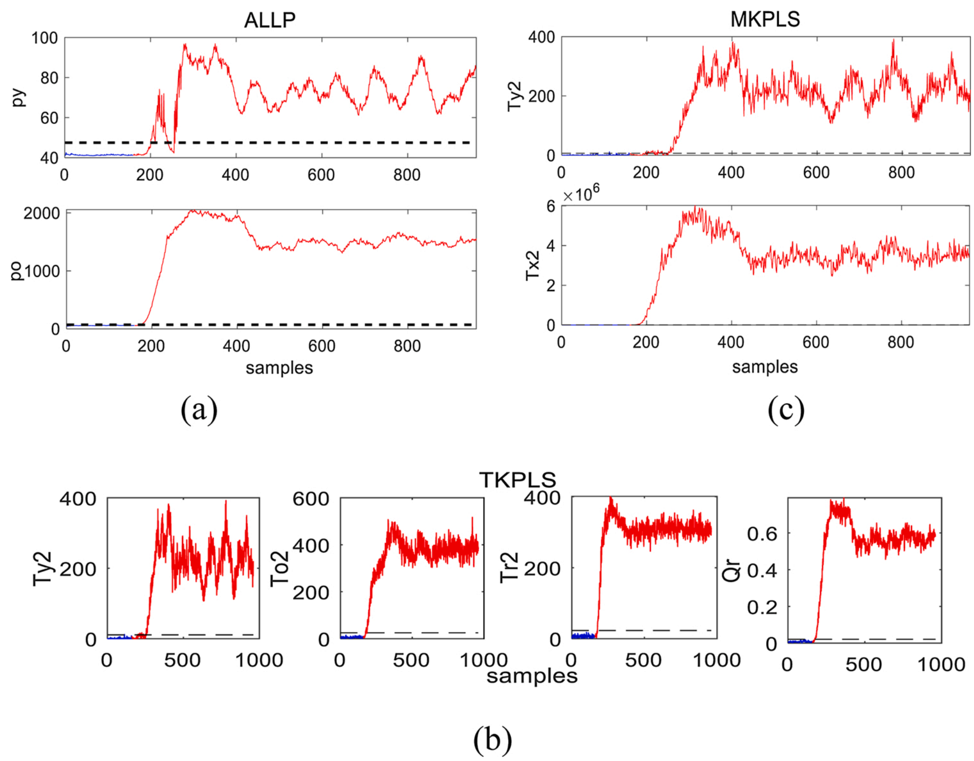

Fig. 8. Monitoring results of the quality-relevant IDV (2) in TEP. (a) Results of proposed method, including statistics p_y and p_o; (b) results of TKPLS, including statistics T_y^2, T_o^2, T_r^2, Q; (c) results of MKPLS, including statistics T_y^2.

variables are modeled by an FCNN with 7−10-1 neurons to predict the values of y. After model construction, the FDRs (%) of the different models are summarized in Table 1. The results show that faults 1 and 2 can be successfully detected through these methods. However, the ALLP can effectively reveal the status of quality variables, especially for the quality-relevant faults. In particular, the monitoring results of fault 1 are presented in Fig. 6.

4.2. Tennessee-Eastman process (TEP)

For quality-relevant fault detection, TEP is an extensively used benchmark, which includes 41 measured variables XMEAS (1)–(41) and 11 manipulated variables XMV (1)–(11). Additional information on this platform can be found in [24,25]. XMEAS (35), which indicates the component of purge gas, is selected as the quality variable. XMEAS (1)–(22) and XMV (1)–(11) are considered to monitor the process, including predicting the values of quality variables. For the convenience of description, $x_1, x_2, \ldots, x_{33}$ and y are used as the process and quality variables in the following description. The datasets sampled for 48 h under normal and faulty conditions are available at http://web.mit.edu/braatzgroup/links.html. A total of 22 datasets are involved, and each set owns 960 samples. The fault-free samples in the training dataset IDV (0) are used to construct the models, and the remaining datasets IDV (1)–(21) in which fault is introduced at the 161 st sample are adopted for testing.

To determine whether a quality-relevant or quality-irrelevant fault occurs, y is assumed to be normally distributed, and the significant level $\alpha = 0.99$ is set to show the variety of quality variables. The quality-irrelevant faults contain IDV (3), (4), (9)–(11), (14)–(17), (19), and (20), thereby indicating that y is unaffected. The quality-relevant faults in the TEP present three statuses. Among them, IDV (1), (5), and (7) are the faults that can be eliminated by the process adjustment. IDV (8), (12), and (13) exhibit oscillating states, and the remaining IDV (2), (6), (18), and (21) indicate the faults that cannot recover to the normal station. For an intuitive representation, the real values of y for IDV (4), (1), (8), and (2) are displayed in Fig. 7, where the normal ranges are

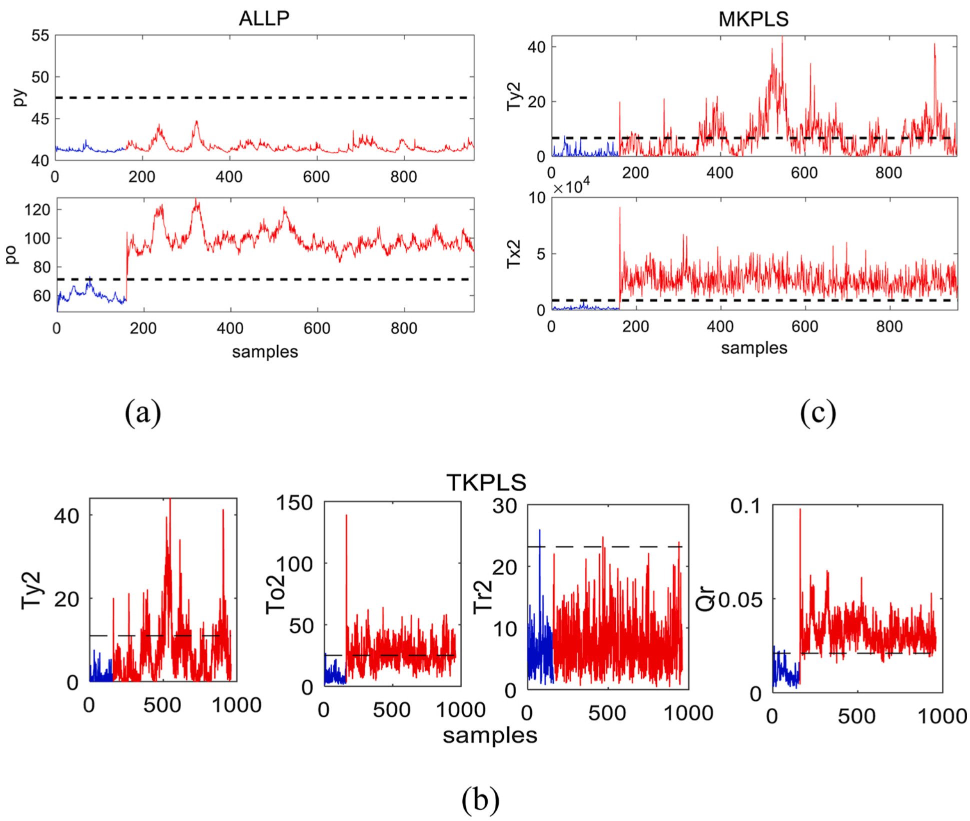

Fig. 9. Monitoring results of the quality-irrelevant IDV (4) in TEP. (a) Results of proposed method, including statistics p_y and p_o; (b) results of TKPLS, including statistics T_y^2, T_o^2, T_r^2, Q; (c) results of MKPLS, including statistics T_y^2, T_x^2.

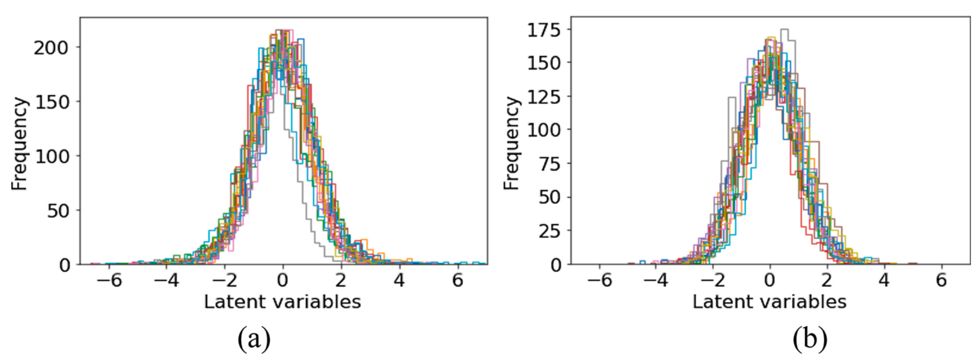

Fig. 10. Frequency distribution histogram of the latent variables extracted using the AAE in two cases of wind turbine blade icing process.

defined between the red dotted lines.

A nonlinear encoder and a linear decoder with 33−40-33 neurons are trained to extract the latent variables. The discriminator and the FCNN, which is used to predict quality variable, are constructed as 40−30-20−1 neurons. The values of y are unaffected during all the faulty times, and n_t, which indicates the rate that y truly deviates from the normal area, is thus calculated. The overall FDRs (%) for quality-relevant and quality-irrelevant faults are summarized in Tables 2 and 3, respectively. Other algorithms, including KPCR, KPLS, TKPLS, and MKPLS, are used for comparison to evaluate the performance of the ALLP, and the optimal performance is highlighted in bold [26]. The average FAR (%) of these methods are similarly small and thus excluded. For most cases, the ALLP outperforms other methods and exhibit competitive monitoring results. The ALLP also provides a reliable result on the quality-relevant statistics based on the real values of y because the FDRs are closest to real values n_t. To reveal the effectiveness of our proposed methods, IDV (2) and (4) are further discussed as follows.

IDV (2) is a step fault that occurs in the B composition with the ratio of A/C constant. The monitoring results using ALLP, TKPLS and MKPLS are given in Fig. 8. IDV (2) is a quality-relevant fault in which quality variable completely deviates from the normal ranges. To be more specific, the real value of quality variable exhibits an oscillated state which is successfully captured by ALLP in Fig. 8(a). The quality-relevant statistic Q of TKPLS does not exhibit this situation in Fig. 8(b) and MKPLS also provide a not competitive result in the quality-relevant statistic in Fig. 8(c). For quality-irrelevant statistics, these methods perform similar results. IDV (4) is a quality-irrelevant fault caused by the step change of condenser cooling water inlet temperature. The monitoring results are provided in Fig. 9. Obviously, the proposed method outperforms other methods in quality-relevant statistics since the false alarms in TKPLS and MKPLS are much too high in Fig. 9(b) and (c). Thus, ALLP can provide a more reliable results especially for reveal the information of quality variables. In general, the superior capability mainly relies on the accurate extraction of quality-relevant features of process variables.

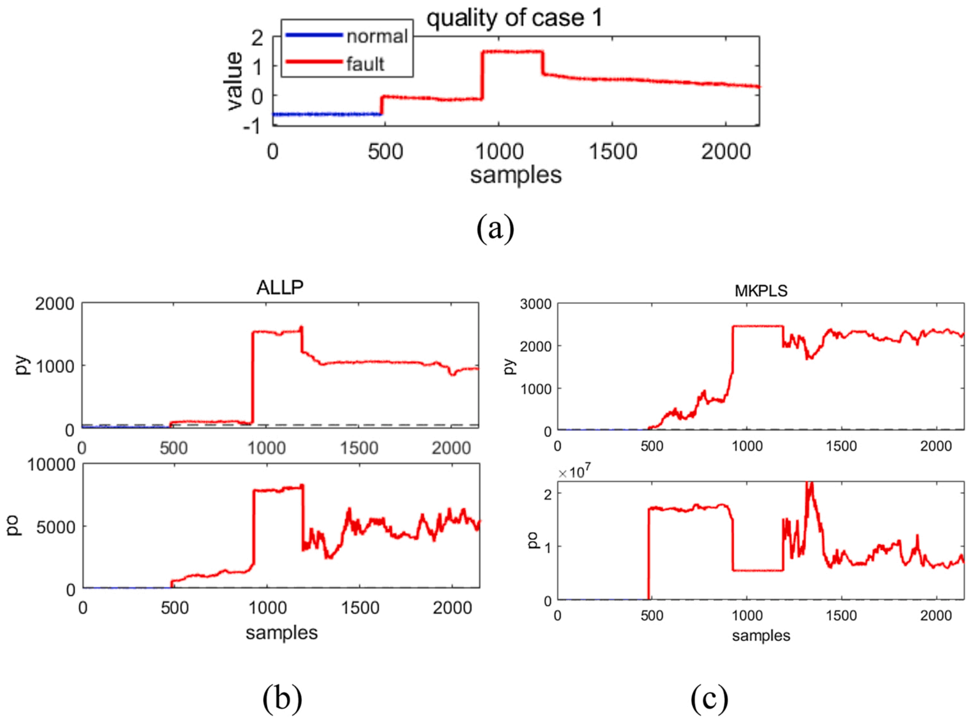

Fig. 11. Monitoring results of case 1. (a) value of quality variable; (b) results of ALLP; (c) results of MKPLS.

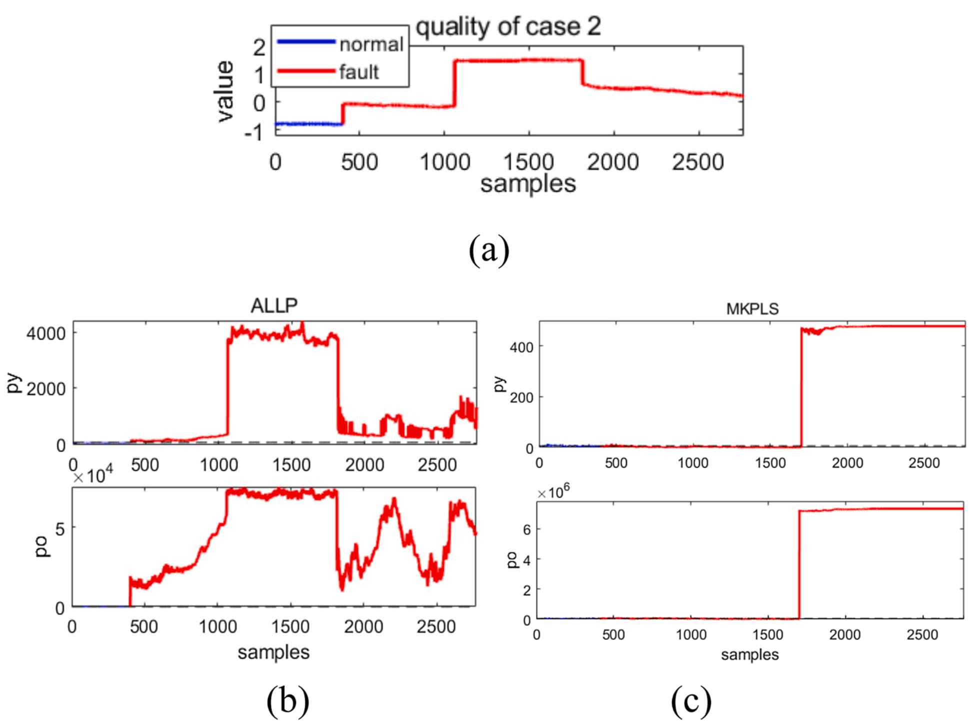

Fig. 12. Monitoring results of case 2. (a) value of quality variable; (b) results of ALLP; (c) results of MKPLS.

5. Application on wind turbine blade icing process

In this section, the proposed scheme is applied in a real industrial process, i.e. wind turbine blade icing process. Blade icing poses a great threat to the power efficiency and safety of the wind turbine and increases the risk of breakage. Currently, most systems can only alarm the severe faults which result in the shutdown of machines and huge economic loss. Therefore, the accurate prediction of the early stage of icing according to the large amount of data recorded every day is challenged and arouses the attention of researchers worldwide.

Following experiment is based on the datasets available at industrial big data industry innovation platform (http://www.industrial-bigdata.com/datasets) in which the descriptions of 28 continuous numerical variables are provided. Among them, the temperature in cabin of wind turbine is chosen as the quality variable since it is key to the safety, and others are selected as process variables expect the recorded time. In the following experiments, two cases involving two wind turbines are studied. One contains 2482 training samples and 2149 testing samples, and other contains 1701 training samples and 2767 testing samples. Besides, the testing datasets consist of a normal stage and three faulty stage.

To perform the proposed scheme, the latent variables of these two cases modeled by adversarial learning are presented in Fig. 10 where we can learn the representative features in the latent space. Given the real values of the quality variables in Figs. 11(a) and 12 (a), blade icing is a quality-relevant fault. Monitoring results of ALLP and MKPLS are respectively provided in Figs. 11 and 12. ALLP performs much better in the false alarms and provides a more accurate prediction of the quality variable. Furthermore, the trends of the predictive information of quality are more accurate by the proposed methods in both cases. In other words, ALLP can reveal the actual states of the quality variables and successfully detect the faults that indicating different degrees of icing. Thus the reliable monitoring results obtained by the proposed method can guide the operators and engineers to a certain extent.

6. Conclusion

In this work, a novel quality relevant fault detection method is studied. This method maps the original variables into a predefined manifold, and then the nonlinear relationship between the latent and quality variables is established. The information of quality relevant contributions can be realized by local projection in the dynamic quality-relevant and quality-irrelevant subspaces under different conditions. In this way, the fault detection is implemented in the two dynamic subspaces in accordance with the probabilities of latent variables. In comparison with traditional projection methods, the proposed method, which uses local projection, exhibits increasingly reliable and efficient monitoring performance for quality-relevant fault detection. Through different case studies, this method can be useful in different scenarios.

The distinguished contribution of latent variables to quality is firstly proposed and it is inspired for the complex data especially for the nonlinearity. And the adversarial manner eases the assumption of assumption about the distribution of raw data and thus it can be widely applied in industrial scenarios. However, the training of models cannot be idealized because real data are consistently incomplete or imbalance. Future works will focus on extracting more representative latent variables for the real-world datasets. To be more specific, how to find an orthogonal information to quality in nonlinear cases remains a challenge.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful for the support of National key research and development program of China (2020YFA0908300), and National Natural Science Foundation of China (21878081).

References

[1] Ge Z, Song Z, Gao F. Review of recent research on data-based process monitoring. Ind Eng Chem Res 2013;52(10):3543–62.
[2] Wang Y, Si Y, Huang B, Lou Z. Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017. Can J Chem Eng 2018;96(10):2073–85.
[3] Xia C, Pan Z, Polden J, Li H, Xu Y, Chen S, et al. A review on wire arc additive manufacturing: monitoring, control and a framework of automated system. J Manuf Syst 2020;57:31–45.
[4] Chen H, Jiang B, Ding SX, Huang B. Data-driven fault diagnosis for traction systems in high-speed trains: a survey, challenges, and perspectives. IEEE trans Intell Transp Syst 2020;21(2):405–65.
[5] Lv Z, Yan X, Jiang Q. Batch process monitoring based on self-adaptive subspace support vector data description. Chemometr Intell Lab Syst 2017;170:25–31.
[6] Leukel J, Gonzalez J, Riekert M. Adoption of machine learning technology for failure prediction in industrial maintenance: a systematic review. J Manuf Syst 2021;61:87–96.
[7] Zhou D, Li G, Qin SJ. Total projection to latent structures for process monitoring. AIChE J 2010;56(1):168–78.
[8] Qin SJ, Zheng Y. Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures. AIChE J 2013;59(2):496–504.
[9] Wang G, Luo H, Peng K. Quality-related fault detection using linear and nonlinear principal component regression. J Franklin Inst 2016;353(10):2159–77.
[10] Li G, Liu B, Qin SJ, Zhou D. Quality relevant data-driven modeling and monitoring of multivariate dynamic processes: the dynamic T-PLS approach. IEEE Trans Neural Netw 2011;22(12):2262–71.
[11] Yin S, Wang G, Gao H. Data-driven process monitoring based on modified orthogonal projections to latent structures. IEEE Trans Control Syst Technol 2016; 24(4):1480–7.
[12] Yin S, Zhu X, Kaynak O. Improved PLS focused on key-performance-Indicator-Related fault diagnosis. IEEE Trans Ind Electron 2015;62(3):1651–8.
[13] Peng K, Zhang K, Li G. Quality-related process monitoring based on total kernel PLS model and its industrial application. Math. Probl. Eng. 2013;2013:707953.
[14] Wang G, Jiao J. A kernel least squares based approach for nonlinear quality-related fault detection. IEEE Trans Ind Electron 2017;64(4):3195–204.
[15] Sun K, Wu W, Liu WT, Yang S, Wang Q, Zhou Q, et al. FAB: a robust facial landmark detection framework for motion-blurred videos. Proceedings of the IEEE International Conference on Computer Vision 2019:5462–71.
[16] Ren L, Sun Y, Cui J, Zhang L. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. J Manuf Syst 2018;48:71–7.
[17] Yu J, Yan X. Active features extracted by deep belief network for process monitoring. ISA Trans 2019;84:247–61.
[18] Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. R J 2016;8(1):289–317.
[19] Ghasemi F, Mehridehnavi A, Fassihi A, Perez-Sanchez H. Deep neural network in QSAR studies using deep belief network. Appl Soft Comput 2018;62:251–8.
[20] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Proceedings of 27th Advances in Neural Information Processing Systems 2014:2672–80.
[21] Makhzani A, Shlens J, Jaitly N, Goodfellow I, Frey B. Adversarial autoencoders. arXiv 2015. arXiv:1511.05644.
[22] Lee J, Lee YC, Kim JT. Fault detection based on one-class deep learning for manufacturing applications limited to an imbalanced dataset. J Manuf Syst 2020; 57:357–66.
[23] Pidhorskyi S, Almohsen R, Adjeroh D, Doretto G. Generative probabilistic novelty detection with adversarial autoencoders. Proceedings of 31st Advances in Neural Information Processing Systems 2018. arXiv: 1807.02588.
[24] Downs JJ, Vogel EF. A plant-wide industrial process control problem. Comput Chem Eng 1993;17(3):145–255.
[25] Ricker NL, Lee JH. Nonlinear model predictive control of the Tennessee Eastman challenge process. Comput Chem Eng 1995;19(9):961–81.
[26] Huang J, Yan X. Quality relevant and independent two block monitoring based on mutual information and KPCA. IEEE Trans Ind Electron 2017;64(8):6518–27.