Research Article

Fault Detection and Health Assessment of Equipment Based on Fuzzy DPCA Spatial Eigenvalue Similarity

Yanhu He,1 Rongyang Wang,1 and Yanfeng Wang2

1College of Mechanical and Electrical Engineering, Huzhou Vocational and Technical College, Huzhou 313000, China
2School of Engineering, Huzhou University, Huzhou 313000, China

Correspondence should be addressed to Yanfeng Wang; neu2009wyf@163.com

Received 3 March 2021; Accepted 23 June 2021; Published 2 July 2021

Academic Editor: Sylvio R. Bistafa

Copyright © 2021 Yanhu He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To improve the fault recognition rate of the dynamic principal component spatial data drive method, a fault diagnosis and equipment health status assessment method based on similarity fuzzy dynamics principal component analysis was proposed. First, the data are fuzzified according to the error function, and an augmented matrix is constructed. The eigenvalues are decomposed to obtain a score matrix and residual matrix of the fuzzy principal component. Further, the similarity between fault data and normal data is calculated. Meanwhile, a health assessment of the equipment is realized. The contribution rate of the observed variables is calculated. Finally, general Tennessee Eastman data and health assessment of a hydraulic press are used to validate the algorithm. The results show that the SFDPCA has a 100% fault recognition rate for some faults, and the recognition rate for other faults is also higher than that of DPCA-Diss, DPCA-SPE, and PCA-SPE. The SDDPCA accurately identifies abnormal phenomena. It can determine the health level of prefilling and effectively make up for the shortcomings of PCA − T², PCA-SPE, DPCA-Diss, and other methods and also can be applied to data-driven fault diagnosis to improve the fault recognition rate.

1. Introduction

To ensure production, it is important to diagnose faults and health assessment of key equipment. Serious consequences would occur, once a fault happens and is not effectively handled [1–4]. Through equipment health assessment, however, minor faults can be found early, facilitating prompt interventions and reducing accidents. At present, methods of fault diagnosis and equipment health assessment mainly include signal analysis methods, model-based methods, and data-driven methods. Signal analysis methods include spectrum analysis, wavelet decomposition (WD), empirical mode decomposition (EMD), and variations of these. Signal analysis methods are not intuitive, are often used for feature extraction, and use model methods for fault diagnosis and health assessment. Model-based methods include support vector machines (SVM), long short-term memory network (LSTMN), neural network (NN), hidden Markov models (HMM), random forests (RF), improved differential evolution [5], expert systems (ES), improved PSO-based QEA [6], and enhanced MSIQDE algorithm [7]. Such methods rely on accurate mathematical modeling and fault data. However, it is difficult to obtain fault data, especially unknown fault data. Reliable industrial process fault data are especially difficult to obtain, making it difficult to use model-based methods for diagnosis [8]. Most complex industrial processes are nonlinear processes, so algorithms such as principal component analysis (PCA), independent component analysis (ICA), and partial least-squares (PLS) have been extensively researched [9]. The basic principle is to monitor normal working data. Once an abnormal change in the observed variables is found, an early warning is given. This is suitable for situations where the model is difficult to build or fault data are difficult to obtain. The DCS and FCS systems are widely used in modern industry, and the data are highly centralized. This provides a good application space for data-driven methods.

The PCA method is one of the data-driven methods. Its core idea is to project the data into the principal component space through linear transformation and make the variance
information of the original data linearly independent. In order to adapt to nonlinear detection, kernel principal component analysis (KPCA) has been proposed. The basic idea is to map low-dimensional nonlinear data to high-dimensional space, convert the nonlinearity of the original data into a linear correlation, and then execute the PCA operation. The main shortcomings of PCA and KPCA are that they cannot adapt to non-Gaussian data and autocorrelations between data [10]. During the work process, as the working conditions change, the statistical characteristics of the variables will change, deviating from the original static standard working conditions, so the dynamic adjustment of data is realized through dynamic principal component analysis (DPCA) [11].

A contribution plot is often used in PCA to identify fault variables [12–15]. PCA usually decomposes the original data into principal component subspace (PCS) and residual subspace (RS), and uses parameters such as the square prediction error (SPE), T², and contribution rate to monitor the status of the data [16]. KPCA and DPCA are currently based on SPE and T² indicators. SPE counts the spatial information changes of the residual space samples, and T² counts the load spatial information changes. SPE and T² indicators can quickly diagnose abnormalities, and they can obtain abnormal variables through contribution plot. In order to improve the recognition accuracy of PCA, some researchers have studied improved PCA algorithms, such as multiblock PCA [17], hierarchical PCA [18], sparse PCA [19], linear time series logic analysis [20], dynamic weight principal component analysis (DWPCA) algorithm [21], and reduced KPCA (RKPCA) approach [22]. In addition, PCA and its improved types are combined with other models to improve the diagnostic accuracy; for example, PCA is combined with IPSOsvm [23], or combined with sliding window principal component analysis (MWPCA) and extreme learning (ELM) [24]. By monitoring the change of eigenvalues in sampling data and normal data, a method of fault diagnosis based on dissimilarity index (Diss) has been proposed [25]. The residual dissimilarity index of PCA-Diss on the basis of Diss has been proposed [26]. The basic idea of Diss and PCA-Diss is to construct a transformation matrix so that the two data spaces have the same eigenvector. The sum of the eigenvalues of the two data spaces is 1 and this can be used for fault detection.

With the improvement of maintenance technology, the health assessment of equipment has been paid more attention. There are two types of PCA-based health assessment algorithms. One is the combination of PCA and its improved methods with models such as recursive principal component analysis (rPCA) and KL divergence fault detection methods [27]. The equipment health assessment of the method is based on the combination of PCA and a least-squares support vector machine [28], the equipment health assessment of the combination of PCA and CMAC [29], and so forth. The other method is to perform eigenvalue decomposition through EMD, wavelets, and so forth and then use PCA-SPE and PCA-T² methods for the equipment health assessment [30].

In summary, SPE and T² indicators are commonly used in PCA for fault detection. The advantage is that the execution speed is fast. But the premise is that the variables obey the multivariate Gaussian distribution and the samples are independent of each other. If this condition is not satisfied, the fault detection rate will be reduced [31].

The Diss method and the PCA-Diss method can solve failures that go undetected by the SPE and T² indicators, and the recognition accuracy is thus improved. However, due to the window scanning method, the execution speed is slow. At the same time, when the eigenvalues between fault data and normal data are similar, misjudgments are common. Moreover, calculating the fault variable is complicated, which affects the accuracy. The equipment health assessment method of rPCA-KL can detect minor faults, but it does not give the relationship between equipment health state changes and divergence, and it is difficult to perform equipment health assessment. Therefore, we propose a fault diagnosis and equipment health assessment method based on similarity fuzzy dynamics principal component analysis (SFDPCA). We design a method of calculating the eigenvalue similarity to monitor the sample space for the purpose of diagnosing faults and health assessment of equipment. The method of calculating the contribution plot is simplified and its effectiveness is verified by experiments.

The main contribution of this paper can be exhibited in the following aspects:

1. The calculation method of eigenvalue similarity in principal component space (SFDPCA) is proposed, and the depth index is designed to select eigenvalues, which can identify different distribution types of data and improve the accuracy of fault identification.

2. A calculation method based on the maximum eigenvalue contribution rate is proposed, which reduces the dependence on the calculation model, reduces the computational complexity, and improves the accuracy.

3. In the principal component space, SFDPCA method is used to evaluate the health status of the equipment, and the health status of the equipment is graded to facilitate engineers to operate and maintain the equipment.

The remaining parts of this article are arranged as follows: Section 2 introduces dynamic principal component analysis, similarity indicators, fault diagnosis, and health status methods. Section 3 introduces health assessment steps based on SFDPCA. Section 4 describes the standard Tennessee Eastman (TE) test and introduces the health status assessment of the hydraulic press prefilling valve. Section 5 offers conclusions.

2. Dynamic Principal Component Analysis and the Eigenvalue Similarity Index

2.1. Fuzzy Dynamic Principal Component Analysis. The fuzzification of samples can effectively improve the recognition rate, so fuzzification is needed first. The process of
fuzzification is a process of weight distribution for each observation variable according to the error. Let the normal sample data be \( X = \{ x_1, x_2, \ldots, x_m \} \), a sample data set with \( m \) observation dimensions. Among them, \( x_i \in \mathbb{R}^n \), where \( x_j \) is the sample observation variable, \( n \) is the number of samples, and the degree of membership \( \zeta = \{ \mu_1, \ldots, \mu_m \} \), \( 0 < \mu_i < 1 \), \( j = 1, 2, \ldots, m \). Then, the fuzzy matrix formed by the sample data can be expressed as

\[
X_\mu = X_\zeta = \begin{bmatrix}
x_{11} \mu_1 & x_{12} \mu_1 & \cdots & x_{1m} \mu_m \\
x_{21} \mu_1 & x_{22} \mu_2 & \cdots & x_{2m} \mu_m \\
\vdots & \vdots & \ddots & \vdots \\
x_{m1} \mu_1 & x_{m2} \mu_2 & \cdots & x_{mm} \mu_m 
\end{bmatrix}.
\] (1)

The objective function of the fuzzy matrix is defined as [32]

\[
J = \sum_{j=1}^{m} \mu_j e(x_j) + \sigma^2 \sum_{j=1}^{m} (\mu_j \log \mu_j - \mu_j).
\] (2)

where \( t \) is the current sampling time. For convenience, let \( X = X_\mu^n \). That is, \( X \) is a matrix with \( n \) rows and \( m(k + 1) \) columns, and subtract the mean value of \( X \) observation variables by column to achieve centralization processing. We then divide the standard deviation of the observation variable by the column to realize the standardization processing:

\[
x_j = \frac{x_j - \bar{x}_j}{\sqrt{\sum_{i=1}^{n} (x_{ij} - \bar{x}_{ij})^2 / n - 1}}.
\] (6)

Here, the fuzzy mean can be expressed as follows:

\[
\bar{x}_j = \frac{\sum_{j=1}^{m} \mu_j x_j}{\sum_{j=1}^{m} \mu_j}, \quad (j = 1, 2, \ldots, m).
\] (7)

Let the standardized fuzzy augmented matrix be \( \bar{X} \), and the covariance \( C \) can be defined as

\[
C = \frac{1}{n-1} \bar{X}^T \bar{X},
\] (8)

where \( C \) is a real symmetric matrix and, thus, can be diagonalized. The diagonal element is the variance, and the nondiagonal element is the covariance. It can be used as an evaluation of the correlation of different components. The smaller the value, the smaller the correlation. The eigenvectors corresponding to different eigenvalues are orthogonal. Let the eigenvalue matrix of \( C \) be \( \Sigma \) and the eigenvector

The error is expressed as

\[
e(x_j) = \left\| x_j - x_j P P^T \right\|^2,
\] (3)

where \( P \) is the eigenvector corresponding to the eigenvalue of the principal component. Then, calculate the partial derivative of \( \mu \). The expression of the membership degree can be obtained as

\[
\mu_j = \exp \left( -\frac{e(x_j)}{\sigma^2} \right),
\] (4)

where \( \sigma^2 \) is the regularization coefficient. Set dynamic delay parameter as \( k \). In this paper, the parallel analysis method is used to determine the value of dynamics delay orders [33]. The fuzzy augmented matrix with dynamic data adjustment characteristics can be generated as follows:

\[
X_\mu^n(t) = \begin{bmatrix}
X_\mu(t) & X_\mu(t-1) & \cdots & X_\mu(t-k) \\
X_\mu(t-1) & \cdots & \cdots & \cdots \\
X_\mu(t-k) & \cdots & \cdots & \cdots \\
X_\mu(t-k-n) & \cdots & \cdots & \cdots
\end{bmatrix},
\] (5)

be \( P \). Then, \( CP = P \Sigma \). We take the diagonal value of \( \Sigma \) to obtain the eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_m (\lambda_1 > \lambda_2 > \cdots > \lambda_m) \) and the corresponding eigenvectors \( p_1, p_2, \ldots, p_m \). Then, PCA fractal dimension algorithm is used to determine the number of principal components [34]:

\[
CDim = \lim_{\epsilon \to \infty} \frac{\log(C(S))}{\log(\epsilon)},
\] (9)

where \( \epsilon \) is the radius of the m-dimensional sphere with \( x_i \) as the center. \( C(S) = \lim_{\epsilon \to \infty} (2/N(N-1)) \sum_{i=1}^{N} \sum_{j>i}^{N} I(\|x_j - x_i\|^2 < \epsilon) \), when \( \|x_j - x_i\|^2 > \epsilon \), \( I = 0 \). When the number of reserved principal components \( r \) is reached, the principal component load matrix \( P = \{ p_1, p_2, \ldots, p_r \} \) and the residual matrix \( \bar{P} = \{ p_{r+1}, p_{r+2}, \ldots, p_{m(k+1)} \} \) can be calculated from the eigenvector. The principal component score matrix is [22] \( T = \bar{x}_{train} \bar{P}^T \), and \( \bar{X} \) can be expressed as

\[
\bar{X} = \bar{x}_{train} \bar{P}^T + e,
\] (10)

where \( e \) is the residual vector.

The SPE and \( T^2 \) indicators are commonly used monitoring parameters, expressed as

\[
T^2 = T \Sigma^{-1} T^T = \bar{x}_{test} \bar{P} \Sigma_{test}^{-1} \bar{P}^T \bar{x}_{test},
\]

\[
SPE = \bar{x}_{test} (I - \bar{P} \bar{P}^T) \bar{x}_{test}^T.
\] (11)
It can be seen that the $T^2$ is a metric for the load space, and the SPE is a metric for the residual space.

### 2.2. Eigenvalue Similarity Index.

After calculating the score matrix, the similarity of the eigenvalues can be calculated. If there are two data $x$ and $y$, the dissimilarity can be expressed as $d = \|x - y\|$; the similarity is expressed as $s = 1/(1 + d)$ and, here, $d$ needs to be normalized so that its value is between 0 and 1 [35]. Let $X$ be the score matrix data set, which has been standardized. It is then divided into two parts, namely, $X = [X_1, X_2]$, where $X_1 = \{x_1^{(1)}, x_2^{(1)}, \ldots, x_{n_1}^{(1)}\}$ contains $n_1$ samples and $X_2 = \{x_1^{(2)}, x_2^{(2)}, \ldots, x_{n_2}^{(2)}\}$ contains $n_2$ samples. $X_1$ is the normal working data set, $X_2$ is the test data set. When $X_1$ is normal data, the similarity threshold can be calculated. $X_2$ is the fault data during fault detection. Both $X_1$ and $X_2$ have been centralized and standardized. Let $M$ be the left covariance matrix of $X$:

$$
M_i = \frac{1}{n_i} - 1 X_i^T X_i, \quad (i = 1, 2).
$$

Then, $M$ is a real symmetric matrix, and there must be $n_i$ eigenvectors orthogonal to each other, which must also be diagonalized. $M$ is eigendecomposed as follows:

$$
M_i Q_i = Q_i \Sigma_i, \quad (i = 1, 2),
$$

where $Q_i$ is a unit orthogonal matrix, which is a matrix composed of the eigenvectors of $M$ that can be used as a new basis vector, $\Sigma_i$ is a diagonal matrix, and the diagonal elements are eigenvalues. Ideally, for the same normal working data set, when the variances of the data in different time periods are the same, $M_1$ and $M_2$ are similar matrices, as follows:

$$
\Sigma_1 = Q_1^{-1} M_1 Q_1 = Q_2^{-1} M_2 Q_2 = \Sigma_2,
$$

where $\Sigma_1 = \{\lambda_1^{(1)} > \lambda_2^{(1)} > \cdots > \lambda_{r_1}^{(1)}\}$ and $\Sigma_2 = \{\lambda_1^{(2)} > \lambda_2^{(2)} > \cdots > \lambda_{r_2}^{(2)}\}$. Using a unified basis vector, we can obtain:

$$
Q_1^{-1} M_1 Q_1 - Q_2^{-1} M_2 Q_2 \leq \xi \epsilon,
$$

where $\xi$ is the unit diagonal vector and $\epsilon$ is the error vector. It can be found that the larger the eigenvalue, the larger $\epsilon$. Let $D$ be the depth coefficient, which represents the number of larger eigenvalues in the eigenvalues sorted in descending order. Let $Q_k^\epsilon$ represent the eigenvector corresponding to the remaining eigenvalues after removing $D$ larger eigenvalues. Then,

$$
(Q_1^\epsilon)^T M_1 Q_1^\epsilon - (Q_2^\epsilon)^T M_2 Q_2^\epsilon \leq \epsilon^2 \xi \approx 0,
$$

$$
\kappa = (D, D + 1, D + 2, \ldots, r),
$$

Equation (16) shows that after unifying the base vector, the similarity between $M_1$ and $M_2$ is related to the selection of the eigenvalue, which is the basis for calculating the similarity. If the expected value and variance of normal data are different after the fault information is injected, this difference will first be reflected in the change of eigenvalues. After the eigenvalues are sorted in descending order, part of the eigenvalue difference just reflects the dissimilarity of the two matrices, namely,

$$
S_{diss} = (Q_1^\epsilon)^T M_1 Q_1^\epsilon - (Q_2^\epsilon)^T M_2 Q_2^\epsilon
\approx \|\text{diag}[\Sigma_1] - \text{diag}[\Sigma_2]\|
\approx \left\|\begin{bmatrix} \lambda_1^{(1)} & \lambda_2^{(1)} & \cdots & \lambda_{D+1}^{(1)} \\ \lambda_1^{(2)} & \lambda_2^{(2)} & \cdots & \lambda_{D+1}^{(2)} \end{bmatrix} - \begin{bmatrix} \lambda_1^{(1)} & \lambda_2^{(1)} & \cdots & \lambda_{D+1}^{(1)} \end{bmatrix} \right\|
\approx \left(\sum_{i = D+1}^{r} \left(\lambda_i^{(1)} - \lambda_i^{(2)}\right)^2\right)^{1/2}.
$$

(17)

The similarity can be calculated as

$$
S = \frac{1}{1 + \sqrt{(1/r - D) \sum_{i = D}^r \left(\lambda_i^{(1)} - \lambda_i^{(2)}\right)^2}}, \quad (i = D, D + 1, \ldots, r),
$$

(18)

where

$$
\lambda_i^{(2)} = \text{diag}\left((Q_1^\epsilon)^T M_2 Q_2^\epsilon\right) = \left(\lambda_1^{(2)}, \lambda_2^{(2)}, \ldots, \lambda_r^{(2)}\right).
$$

(19)

Equation (19) expresses the similarity of the two matrices. At the same time, it is found that the trace of the matrix can reflect the change of the fault, so the similarity can be optimized as follows:

$$
S = \frac{1}{1 + \sqrt{(1/r - D) \sum_{i = D}^r \left(\lambda_i^{(1)} - \lambda_i^{(2)}\right)^2}} \eta + \text{trace}(\Sigma_2)(1 - \eta),
$$

(20)

where $\eta$ is the adjustment factor, which represents the weight distribution between the two parts.

It should be noted that the value of $D$ cannot exceed the number of eigenvalues, and the number of eigenvalues is the value $r$ after the dimensionality reduction of the augmented matrix.

### 2.3. Fault Diagnosis.

When an abnormality occurs, it is necessary to find the variable causing the fault and determine the fault source. The determination of the fault variable is generally completed through a contribution plot. Let $\Theta$ be a set of data samples with $w$ faulty data that have been sampled. First, we use the mean and variance of normal data for normalization. From (12) and (13), we can get

$$
\frac{1}{n_i} - 1 X_i^T X_i Q_1 = Q_i \Sigma_i.
$$

(21)

The score matrix of the failure data is
Mathematical Problems in Engineering

\[ T_{Q} = \Theta_{Q}^T, \]  \tag{22} 

where \( \Theta_{Q} \) represents the corresponding eigenvector when the principal element is \( r \) after dimensionality reduction; it can be concluded that

\[ \frac{1}{w - 1} \Theta_{Q}^T \Theta_{Q} = Q_{j}^T \Sigma_{2}. \]  \tag{23} 

Then, in the \( T_{Q} \) sliding window, eigenvalue decomposition is performed on \( T_{Q} \). The eigenvalue and eigenvector are calculated with (24), and the score matrix is calculated with (25):

\[ \frac{1}{w - 1} T_{Q}^T T_{Q} P_{w} = P_{w} \Sigma_{w}, \]  \tag{24} 

\[ T_{Q} = T_{Q} P_{w}. \]  \tag{25} 

When calculating the contribution degree according to the traditional method, it needs to be calculated in the space given by (24) and (25), because after twice calculating the variance and eigenvalue decomposition, the fault variable and eigenvalue contribution rate model are difficult to determine and are complicated. According to matrix theory, it can be obtained that

\[ \frac{1}{w - 1} T_{Q}^T \Theta_{Q}^T \Theta_{Q} P_{w} T_{Q} = \frac{1}{w - 1} \Theta_{Q}^T \Theta_{Q} P_{w} \sum_{i} \lambda_{i} = \sum_{i} \lambda_{i} P_{w} = P_{w} \sum_{i} \lambda_{i}. \]  \tag{26} 

It can be deduced that

\[ \Sigma_{w} = P_{w} \sum_{i} \lambda_{i} P_{w}. \]  \tag{27} 

Therefore, \( \sum_{w} \) and \( \sum_{2} \) are similarity matrices, so the contribution degree only needs to be calculated in the space where \( \sum_{2} \) is located. This greatly reduces the computational complexity. Further, it can be found that the most obvious change in the eigenvalues is the maximum eigenvalue. Therefore, the contribution degree can be calculated according to the degree of influence of the maximum eigenvalue. The formula is as follows:

\[ \Gamma_{k} = p^T \Theta^T \Theta, \]  \tag{28} 

\[ \lambda(i) = \frac{\Gamma_{k} (i) p(i)}{\Gamma_{k} p}, \]

where \( p \) is the eigenvector corresponding to the largest eigenvalue.

2.4. Health Status Assessment. The health assessment of the equipment is expressed by similarity. When a component of the equipment has fault symptoms, it can be observed by the change of observation variables, and the observation variables deviating from the normal value can be accurately measured by the degree of equipment performance degradation. According to the deviation of the fault variable from the normal value, the health status is divided into four levels. It is generally divided into health, subhealth, fault alarm, and failure. The failure sample set after standardization is

\[ \overline{x}_{k} = \{ x_{j1}^{k}, x_{j2}^{k}, \ldots, x_{jm}^{k} \}, \]  \tag{29} 

where \( j \) is the health assessment level, \( \overline{x}_{j}^{k} \) is an abnormal observation variable at level \( j \), \( x_{0}^{k} \) indicates a normal value, and \( x_{m}^{j} \) indicates a value changed at level \( j \). At this time, the similarity calculation formula is as follows:

\[ S_{ij}^{j} = \frac{1}{1 + \sqrt{(1/r - D) \sum_{j} (\overline{P}_{w}^T \overline{P}_{j}) (\overline{P}_{w}^T \bar{P}_{j} - \bar{P}_{w}^T \bar{P}_{j}^{-1})}} \eta + \text{trace} \left( \sum_{2} \right) (1 - \eta). \]  \tag{30} 

where \( n \) is the number of samples in state \( j \).

3. Health Status Assessment Steps Based on SFDPCA

According to the SFDPCA eigenvalue similarity calculation method, the equipment fault is detected and the health status assessment is carried out. The steps are as follows:

Step 1: Two groups of normal working data \( X_{1} \) and \( X_{2} \) are collected and fuzzified, and then the augmented matrix is constructed according to the amplification coefficient, which is recorded as \( \bar{X}_{1} \) and \( \bar{X}_{2} \).

Step 2: Use the mean and variance of \( \bar{X}_{1} \) to centralize and standardize \( \bar{X}_{1} \) and \( \bar{X}_{2} \), which are recorded as \( \bar{X}_{1} \) and \( \bar{X}_{2} \).
Step 3: The covariance matrix of \( \bar{X}_1 \) is calculated as \( C_1 \). Calculate the eigenvalue and eigenvector of \( C_1 \). Set the number of principal elements as \( r \), and calculate the load matrix \( P_r^{(1)} \).

Step 4: Use \( P_r^{(1)} \) to calculate the score matrix \( T_1 \) and \( T_2 \). The mean and variance of \( T_1 \) are used to centralize and standardize \( T_1(i = 1, 2) \), namely, \( \bar{T}_1(i = 1, 2) \).

Step 5: Calculate the eigenvalue and eigenvector of \( T_1 \). At the same time, select window data \( D_w^o \) of window width \( w \) on \( T_2 \). Slide the window according to the step size \( s \). Calculate the eigenvector and eigenvalue of \( D_w^o \), and find the control limit of similarity as \( S_m \).

Step 6: Collect a set of window width data, calculate the fuzzy matrix, and then construct its augmented matrix of \( X \). The mean and variance of \( X_1 \) are used to standardize \( \Theta_{(1)} \) and record it as \( \Theta_{(1)}^{+} \). Then, \( P_r^{(1)} \) is used to calculate the score matrix \( \bar{T}_{(1)}^{+} \), and then the similarity between \( \bar{T}_{(1)}^{+} \) and \( \bar{T}_1 \) is calculated. If it is greater than the control limit, it is a fault.

Step 7: Once a failure occurs, the fault variable is determined according to the contribution value.

Step 8: Determine the health status level according to the curve of the similarity.

Remark. The control limit can also be determined by the empirical method of the 99% confidence upper limit [36]. 99% confidence level is obtained from 99% test sample statistics. It is worth noting that the empirical method is used for this paper.

4. Results and Analysis

4.1. Test with Standard Cases. TE is a process simulator that is widely used to verify fault diagnosis algorithms. The TE process consists of five main components: a reactor, a condenser, a vapor-liquid separator, a circulating compressor, and a product stripper. The TE process includes four gas raw materials \( A, D, E, C \), and two liquid products \( G, H \), which are produced by reaction downstream. This describes a chemical reaction process that includes 12 operating variables, 22 continuous process measurement variables, and 19 component measurement variables. TE provides 21 sets of standard training and test data sets. The first group of data comprises normal data, and the remaining 20 groups are fault data. Each group of data includes training data and test data. This section uses the TE test data set, which has 960 samples. Each sample has 52 observed variables. It should be noted that, in the fault data, the first 160 data samples are normal data, and faults are introduced thereafter.

4.2. Recognition Rate Test. In the TE fault set, the No. 9 fault is a random fault, and the No. 19 fault is an unknown fault. The results measured by traditional PCA-SPE, PCA-\( T^2 \), DPCA-SPE, and other methods are not ideal. By contrast, the proposed SFDPCA method has a recognition rate of 99.8% for the No. 9 fault, and the misjudgment rate is 0, as shown in Figure 1. The recognition rate of the No. 19 fault reaches 100%, and the fault misjudgment rate is 0, as shown in Figure 2. The parameters used are \( k = 3, w = 60, D = 2 \) and the weight factor is 0.7.

It can be seen from Figures 1 and 2 that before the 160 data points of the fault introduction point, the data measured by the two fault SFDPCA methods are all below the control line, indicating that normal data can be recognized normally. However, data fluctuations measured by the PCA-SPE and PCA-\( T^2 \) methods higher than the control line indicate recognition errors. After 160 data points, the SFDPCA method recognizes the No. 19 fault. The data after the fault occurs are all higher than the control line, indicating that the recognition is correct. Thus, the recognition rate reaches 100%. For the No. 9 fault, there is only one data recognition error, and the recognition rate reaches 99.8%. Whether it is the No. 9 or No. 19 fault, many of the SPE and \( T^2 \) data are under the control limit, and the recognition rate is low. The test with other TE data achieves good results. The comparison between SFDPCA and other methods is shown in Table 1.

The above standard test shows the effectiveness of this algorithm.

According to fault test results of TE data, for different fault types, the test results of PCA-SPE, PCA-\( T^2 \), DPCA-SPE, Diss, and DPCA-Diss are very different. That is, for the same fault, some methods have a high recognition rate, while others have a low recognition rate. At this time, because different methods have different adaptability to data distribution types (such as Gaussian distribution), the results of SFDPCA are higher than those of PCA-SPE, PCA-\( T^2 \), DPCA-SPE, Diss, and DPCA-Diss, which indicates that SFDPCA is robust to data distribution. This is because SFDPCA adopts the depth index method, cancels the large eigenvalue to participate in the similarity calculation, and avoids the impact of large eigenvalue on small eigenvalue to affect the test results. At the same time, SFDPCA adopts the dynamic constant method and fuzzy factor to suppress the observation variables that have a great impact on the recognition results, and improves the recognition rate, which is different from the DPCA-Diss method, and SFDPCA considers the influence.

According to equation (28), the contribution plot is used to determine TE fault variables. Now take No. 10 fault as an example. The TE process shows that the abnormal change of observed variable 18 is the main cause of fault 5. And its contribution plot is shown in Figure 3. It can be seen from Figure 3 that the stripper temperature (variable 18) has an abnormal fluctuation, and the algorithm results are consistent with the actual situation.

4.3. Depth Index Test. The depth index \( D \) has a great impact on similarity. The depth index test is carried out with the No. 15 fault, and the results are shown in Figures 4–6.

It can be seen from Figure 4 that when \( D = 1 \), the similarity fluctuates up and down on the control line. Most of the 160 signals before the introduction of the fault appear above the control line and are mistakenly divided into fault signals (i.e., the normal signal is below the control line). The
signal after the introduction point of the fault signal (160) also has a large number of identification errors (i.e., the fault signal should be above the control line).

When $D = 5$, the recognition rate of the fault signal is improved, but the normal signal is still above the control line, and most of the signals are wrong.

When $D = 10$, normal signals can be correctly identified, most of the fault signals can be correctly identified, and the recognition rate is greatly improved. The above test shows that the depth index has a greater impact on the similarity, and the correct choice has an important effect on signal recognition.

4.4. Fuzziness Effect Test. In order to test the impact of fuzzification on similarity, each observation variable is fuzzified. Then, the test is conducted with $D = 10$, as shown in Figure 7.
It can be seen from Figure 7 that when the fuzzy algorithm is used, the 160 normal signals before the introduction of the fault point are all under the control line, indicating that they have been correctly identified. At the same time, the fault signals after the introduction of the fault point (160) are all above the control line. The fault signal is thus fully recognized.

4.5. Dynamic Index Test. The choice of the time delay constant also affects the recognition rate. We use the No. 19 fault as an example for testing, as shown in Figures 8 and 9. It can be seen from Figure 8 that when the dynamic index is one, that is, when the augmented matrix is not used, the recognition rate of the normal signal at 160 points before the fault introduction point is high, but the fault data after the introduction of the fault point have errors. The error is between 400 and 500.

It can be seen from Figure 9 that when the time lag constant increases, the recognition rate is significantly improved, indicating that the recognition rate improves after adding dynamic characteristics. In practice, considering the execution efficiency, the time delay constant should not be high.

4.6. Health Status Assessment of Hydraulic Press Prefilling Valve. The hydraulic press is an indispensable device in modern manufacturing. The prefilling valve is an important hydraulic component in hydraulic presses. Its main function is to overcome high pressure and open the pressurized equipment through the control requirements, for rapid pressure relief and large flow of hydraulic oil return tank. The prefilling valve is a common component in hydraulic presses, and its monitoring is of great significance.
4.6.1. Hydraulic Model of Prefilling Valve. The structure of the prefilling valve is shown in Figure 10. It is composed of the control oil circuit, oil cup, main valve core, and other components.

The degradation of the prefilling valve health status mainly manifests that the main valve cannot be opened. According to the hydraulic balance equation, the opening pressure model can be expressed as follows:

\[ P \Delta A_{eq} - F_s - F_{\mu} - G = M \frac{\Delta v}{\Delta t} \eta, \]

\[ P = P_1 - \Delta P_w - \Delta P_z - \Delta P_r, \]

where \( P_1 \) is the control pressure; \( \Delta P_w \) is the pressure loss along the pipeline, which is the pressure loss when the hydraulic pressure flows in the pipeline and is related to the
hydraulic flow rate; \( \Delta P\) is the local pressure loss, which is the pressure loss of the hydraulic pressure in the control oil cup; \( \Delta P_r\) is the pressure loss caused by prefilling valve leakage; \( A_1\) is the piston area of the control cylinder; \( P_2\) is the hydraulic pressure outside the main spool; and \( A_{eq}\) is the main valve core area. The shape of the main valve core varies, and here it refers to the effective area after conversion. Further, \( F_s\) is the spool spring pressure; \( F_\mu\) is friction resistance; \( G\) and \( M\) are the weight and mass of the spool group, respectively; \( \Delta v/\Delta t\) is the spool start acceleration.
Figure 8: Similarity when $k = 1$ and $D = 5$.

Figure 9: Similarity when $k = 3$ and $D = 5$.

Figure 10: Structure diagram of the precharge valve.
index; and η denotes efficiency, which is a comprehensive measure of pressure loss caused by heating and other reasons, with a value of 0.9–0.95:

\[ \Delta P_w = \sum \lambda \frac{l \rho v^2}{2}, \tag{34} \]

\[ \Delta P_\varsigma = \sum \zeta \frac{\rho v^2}{2}, \tag{35} \]

where \( \lambda \) is the drag coefficient along the way, \( \lambda = 0.028 \); \( l \) is the pipeline length; \( d \) is the inner diameter of the pipeline; \( \rho \) denotes oil density; \( v \) is the average flow rate of the pipeline; \( \zeta \) is the local resistance coefficient. According to the thick wall flow formula, the approximate formula of the gap leakage can be obtained:

\[ \Delta P_r = \frac{\rho v^2}{2C_q}, \tag{35} \]

where \( C_q \) is the flow coefficient, which can be determined through experiments. Then, there is

\[ P = P_1 - \left( \sum \lambda \frac{l \rho v^2}{2} + \sum \zeta \frac{\rho v^2}{2} + \frac{\rho v^2}{2C_q} \right)v^2. \tag{36} \]

It can be seen from (36) that the pressure loss is directly proportional to the square of the speed. When the oil cup does not leak, the pressure of the oil is in a closed state in the oil cup, and the pressure loss can be ignored. At this time, \( P = P_1 \). When a gap in the oil cup begins to leak, the pressure loss caused by the leak increases with an increase in the liquid flow rate. Then, \( P \) begins to decrease and decreases rapidly with an increase in leakage. When \( P \) drops to the critical point \( P_0 \),

\[ P_0A_1 - P_2A_{eq} - F_s - F_\mu - G < M \frac{\Delta v}{\Delta t} \frac{1}{\eta}. \tag{37} \]

This indicates that the prefilling valve is damaged, unable to open, and unable to relieve pressure.

4.6.2. Equipment Health Status Assessment and Fault Diagnosis of Prefilling Valve Based on SFDPCA. The experimental object is a hydraulic press used for short-period
veneer production. The NI9203 pressure and flow acquisition card is used for data collection. The input signal is buffered, conditioned, and sampled by 16 bits of ADC. CDAQ9189 is adopted as a signal server, which mainly completes the conversion and integration of field signal transmission protocol and uploads it to the host computer in the form of Ethernet. The upper computer software is LabVIEW.

The prefilling valve for this hydraulic press does not use a pilot device, so a larger opening pressure is required. The control protection pressure is 14 MPa, and the normal opening pressure ranges from 3 MPa to 11 MPa. A pressure-reducing valve is designed on the control oil circuit of the prefilling valve so that the pressure loss of the prefilling valve does not affect the work of other valves.

According to the opening pressure drop, it is divided into four levels to illustrate the process of valve damage: the healthy level, the subhealthy level, the fault alarm level, and the fault level. According to (30), the similarity under the health assessment level can be calculated, as shown in Table 2.

We use 960 samples of working press data. Each sample includes nine observation variables, that is, hydraulic lock opening pressure, lifting cylinder lifting pressure one, lifting cylinder lifting pressure two, press falling pressure, prefilling valve opening pressure, system working pressure, pressurizing pressure, holding pressure, and pressure relief pressure. The similarity is used to indicate the trend of health assessment changes, as shown in Figure 11.

It can be seen from Figure 11 that, before sampling point 200, the press is healthy and the similarity does not change significantly. After sampling point 200, the similarity begins to change and reaches the subhealthy level after 480 points. Although the system can work, a warning should be issued. After 670 points, the system shows a tendency of accelerated change and enters the fault alarm level, indicating that the valve has a serious fault. Although it can work, it should be repaired. When the similarity reaches the fault level, the system cannot work.

Degradation rate is often used to observe the degree of degradation. According to the degradation rate, the health status is divided as shown in Figure 12.

The numerical relationship between health status and degradation rate is shown in Table 3.

In practice, health status changes are caused by one or more variables. In order to accurately locate abnormal fault variables, it is necessary to calculate the contribution of each observed variable to the maximum eigenvalue because the abnormality of the fault variable will first cause the change of the largest eigenvalue. According to this principle, we can observe the contribution of each variable through the largest
eigenvalue. As shown in Figure 13, the contribution of the first 300 sampling points is basically the same. After 300 sampling points, the contribution shown in red increases greatly, and the corresponding variable is the fifth variable.

A histogram can be used to illustrate the contribution of a certain sampled data observation variable. Figure 14 shows the contribution of the 700th sampled data after the failure. The contribution of variable 5 is significantly higher than...
other variables, indicating that the source of the failure is this variable.

In order to study the change trend of the eigenvalues of each window in the score matrix and the degree of impact on the similarity, the eigenvalues of each window are analyzed to obtain their change trends, as shown in Figure 15.

It can be seen from Figure 15 that after sampling point 200, each eigenvalue begins to change. The eigenvalues

**Figure 15:** Change trend of eigenvalues of each window.

**Figure 16:** Eigenvalues of normal data and fault data.
gradually increase, indicating that abnormal data is generated. It can be seen that when a fault occurs, the response of eigenvalue 1 (the maximum eigenvalue) is the most obvious. The changes in other eigenvalues decrease sequentially. Therefore, the contribution degree of each observed variable can be calculated through the contribution degree of the maximum eigenvalue. Then, the fault occurrence variable can be determined. However, the similarity is often calculated by the eigenvalues with small changes.

To study the relationship between normal eigenvalues and fault eigenvalues, 10 normal samples (1–10) and 10 fault samples (300–310) are selected for observation, as shown in Figure 16.

It can be seen from Figure 16 that after a fault occurs, the eigenvalues are significantly different from those of normal data. As the fault escalates, the eigenvalues show an increasing trend. It is thus easy to calculate the trends of change and size by using similarity.

5. Conclusions
The SFDPCA method proposed in this paper is suitable for the monitoring of equipment working, and its ability to identify abnormal fault data is better than DPCA-SPE, PCA-T², PCA-SPE, Diss, DPCA-Diss, and other methods. The test results on TE data show that the SFDPCA has a higher recognition rate for TE fault signals, and the recognition rate for fault 2 and fault 6 of TE data reaches 100%. Compared with DPCA-SPE, PCA-T², PCA-SPE, Diss, and DPCA-Diss methods, SFDPCA reduces the computational complexity and improves the accuracy. At the same time, SFDPCA divides the similarity into four levels according to the fault variables, realizes the monitoring of equipment health status, and provides a scientific basis for the maintenance, routine maintenance, and spare parts management of large equipment. Through the contribution diagram and decomposition diagram of the similarity of each variable, the change of fault variable can be clearly observed, which provides a reference for engineering and technical personnel to carry out fault diagnosis. In the data test of the hydraulic press precharging valve, the abnormal observation variable can be accurately identified, and the abnormal level is given, which shows the effectiveness of the algorithm. SFDPCA needs to adjust the depth index, ambiguity, window width, dynamic constant, degradation factor, and other parameters to adapt to different conditions. In the future research, further optimize the parameters and realize adaptive adjustment.

Data Availability
Readers can contact the corresponding author to obtain the relevant data for this study.

Conflicts of Interest
The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments
This research was supported by the Zhejiang Public Welfare Technology Research Project (LGG19E050005), Zhejiang Natural Science Foundation of China (LY19F030001), Huzhou Science and Technology Plan Project (2020GG33), Science and Technology Program of Zhejiang Province Key Laboratory of Smart Management, General Research Project of Zhejiang Provincial Department of Education (Y202044896), and Application of Modern Agricultural Resources (2020E10017).

References
[1] Y. Wang, P. He, P. Shi, and H. Zhang, “Fault detection for systems with model uncertainty and disturbance via coprime factorization and gap metric,” IEEE Transactions on Cybernetics, 2021.
[2] Z. Gao, C. Cecati, and S. X. Ding, “A survey of fault diagnosis and fault-tolerant techniques—Part I: fault diagnosis with model-based and signal-based approaches,” IEEE Transactions on Industrial Electronics, vol. 62, no. 6, pp. 3757–3767, 2015.
[3] Y.-F. Wang, “Robust H∞ fault detection for networked control systems with markov time-delays and data packet loss in both S/C and C/A channels,” Mathematical Problems in Engineering, vol. 2019, Article ID 4672862, 11 pages, 2019.
[4] Z. Li, P. Wang, Z. Zhou et al., “Robust fault detection for networked markov jump systems with random time-delay,” Mathematical Problems in Engineering, vol. 2017, Article ID 2608140, 9 pages, 2017.
[5] W. Deng, S. Shang, X. Cai et al., “Quantum differential evolution with cooperative coevolution framework and hybrid mutation strategy for large scale optimization,” Knowledge-Based Systems, vol. 224, Article ID 107080, 2021.
[6] W. Deng, J. Xu, H. Zhao, and Y. Song, “A novel gate resource allocation method using improved PSO-based QEA,” IEEE Transactions on Intelligent Transportation Systems, pp. 1–9, 2020.
[7] W. Deng, J. Xu, X.-Z. Gao, and H. Zhao, “An enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, pp. 1–10, 2020.
[8] Q. Liu, J. Zhuo, Z. Lang, and S. Qin, “Perspectives on data-driven operation monitoring and self-optimization of industrial processes,” Acta Automatica Sinica, vol. 44, no. 11, pp. 1944–1956, 2018.
[9] Q. Zhu, Q. Liu, and S. J. Qin, “Quality-relevant fault detection of nonlinear processes based on kernel concurrent canonical correlation analysis,” in Proceedings of the 2017 American Control Conference (ACC), pp. 5404–5409, IEEE, New Orleans, LA, USA, 2017.
[10] K. Peng, L. Ma, and K. Zhang, “Review of quality-related fault detection and diagnosis techniques for complex industrial processes,” Acta Automatica Sinica, vol. 43, no. 3, pp. 349–365, 2017.
[11] W. Ku, R. H. Storer, and C. Georgakis, “Disturbance detection and isolation by dynamic principal component analysis,” Chemometrics and Intelligent Laboratory Systems, vol. 30, no. 1, pp. 179–196, 1995.
[12] J. A. Westerhuis, S. P. Gurden, and A. K. Smilde, “Generalized contribution plots in multivariate statistical process
monitoring,” *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 95–114, 2000.

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

[14] M. Kano, S. Hasebe, I. Hashimoto, and H. Ohno, “Evolution of multivariate statistical process control: application of independent component analysis and external analysis,” *Computers & Chemical Engineering*, vol. 28, no. 6-7, pp. 1157–1166, 2004.

[15] C. Zhao and F. Gao, “Fault subspace selection approach combined with analysis of relative changes for reconstruction modeling and multia fault diagnosis,” *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 928–939, 2015.

[16] C. Zhang, Q. Guo, L. Feng, and Y. Li, “Fault detection strategy based on locality preserving projections-weighted k nearest neighbors in multimodal batch processes,” *IEEE Transactions on Control Systems Technology*, vol. 36, no. 10, pp. 1682–1689, 2019.

[17] Q. Jiang, X. Yan, and B. Huang, “Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference,” *IEEE Transactions on Industrial Informatics*, vol. 63, no. 1, pp. 377–386, 2015.

[18] S. Wold, N. Kettaneh, and K. Tjessem, “Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection,” *Journal of Chemometrics*, vol. 10, no. 6, pp. 463–482, 1996.

[19] Y. Liu, G. Zhang, and B. Xu, “Compressive sparse principal component analysis for process supervisory monitoring and fault detection,” *Journal of Process Control*, vol. 50, pp. 1–10, 2017.

[20] B. Di Giampaolo, S. La Torre, and M. Napoli, “Parametric metric interval temporal logic,” *Theoretical Computer Science*, vol. 564, pp. 131–148, 2015.

[21] Y. Tao, H. Shi, B. Song, and S. Tan, “A novel dynamic weight principal component analysis method and hierarchical monitoring strategy for process fault detection and diagnosis,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 7994–8004, 2019.

[22] F. Bencheikh, M. F. Harkat, A. Koudadi, and A. Bensmail, “New reduced kernel PCA for fault detection and diagnosis in cement rotary kiln,” *Chemometrics and Intelligent Laboratory Systems*, vol. 204, Article ID 104091, 2020.

[23] T. Yi, Y. Xie, H. Zhang, and X. Kong, “Insulation fault diagnosis of disconnecting switches based on wavelet packet transform and PCA-IPSO-SVM of electric fields,” *IEEE Access*, vol. 8, pp. 176676–176690, 2020.

[24] K. Song, P. Xu, Y. Chen, T. Zhang, G. Wei, and Q. Wang, “A fault diagnosis and reconconfiguration strategy for self-validating hydrogen sensor array based on MWPCA and ELM,” *IEEE Access*, vol. 7, pp. 115075–115092, 2019.

[25] G. Wang, J. Liu, and Y. Li, “A progressive fault detection and diagnosis method based on dissimilarity of process data,” in *Proceedings of the 2014 IEEE International Conference on Information and Automation (ICIA)*, pp. 1211–1216, Hailar, China, 2014.

[26] C. Zhang, X. Dai, and Y. Li, “Fault detection and diagnosis method based on DPCA residual disparity,” *Acta Automatica Sinica*, pp. 1–10, 2020.

[27] W. Zhou, H. Pan, and P. Wu, “Small fault detection method based on DPCA and KL divergence,” *Transducer and Microsystem Technologies*, vol. 39, no. 3, pp. 135–138, 2020.