Contents lists available at YXpublications

International Journal of Applied Mathematics in Control Engineering

Journal homepage: www. http://www.ijamce.com

KNN Transition Period Fault Detection Based on Random Projection

Mengya Shao[a], Feng Lv[b*], Wenxia Du[b], Xuejun Song[a]

[a]*College Of Physics And Information Engineering, Hebei Normal University, Shijiazhuang, 050000, China;*

[b]*College Of Career Technology, Hebei Normal University, Shijiazhuang, 050000, China.*

ARTICLE INFO

Article history:
Received 23 July 2018
Accepted 12 October 2018
Available online 25 December 2018

Keywords:
batch process
MPCA
FD-KNN
Random projection
KNN
fault diagnosis

ABSTRACT

Intermittent process has the characteristics of short response period, fast internal dynamic change, and interrelated variables. The transition period data is more complex, and the data characteristics change quickly. The traditional MPCA (Multivariate Principal Component Analysis) fault diagnosis method uses PCA (Principal Component Analysis) to reduce dimensionality, which not only does not save the distance information of data, but also has a limited scope of use. FD-KNN (Fault Detection based on kNN) can not reduce the dimension of data, which results in too much computation and increases the time of fault detection. Based on this and in this paper, a KNN (k-Nearest Neighbor Rule) transition time fault detection method based on random projection is introduced. In this method, the transition modal data is projected from high-dimensional space to low-dimensional space by a de-randomized JL (Johnson-Lindenstrauss) transformation, and the projection matrix is quickly generated when the distance between any two points is protected. Then the fault diagnosis is carried out by KNN method. The simulation results on machining data from machine tools show that the method improves the accuracy of fault diagnosis and enlarges the scope of fault diagnosis.

1.Introduction

The methods of fault diagnosis in the industry are endless, and multivariate statistical analysis and machine learning in data-driven methods have always been hot research objects.

Multivariate statistical analysis uses the correlation between multiple variables collected by sensors to diagnose faults. Monitor based on whether online statistic exceeds the range of normal data established before. The commonly used multivariate statistical analysis methods have Principal Component Analysis(PCA), Partial Least Squares(PLS), Independent Component Analysis(ICA) and so on [1]. Bakshi introduced the wavelet transform into PCA and proposed the MPCA (Multiscale PCA, MsPCA) method [2]. Other methods for PCA extension include Adaptive Principal Component Analysis [3], Nonlinear Principal Component Analysis [4], etc. With the deepening of data analysis, the division and diagnosis methods of intermittent processes are also constantly improving. Yao [5] and Zhao et al. [6], hardening and softening the data separately and monitoring and quality prediction based on the intermittent characteristics of the time period. Although the PCA-based fault diagnosis method is simple, easy to calculate, and there is some result in fault detection, but it is difficult to clearly explain the cause of the fault [7], and the method needs to assume that the variable obeys the multivariate normal distribution. In the face of linearity and nonlinearity, Gaussian and non-Gaussian problems need to establish different models. A machine learning method can provide a good solution to this type of problem.

The machine learning class approach is to treat the fault diagnosis problem as a classification problem. He et al. [8] proposed a fault detection algorithm based on k-nearest neighbor (KNN) to solve Gaussian non-Gaussian problem. Ma Hehe et al. [9] proposed a local outlier factor (LOF) method based on the Mahalanobis distance to detect faults, and realized the purpose of monitoring the multimodal process by establishing a single model. Guo Jinyu et al. [10] proposed a batch process KNN fault detection method based on online upgrade master sample modeling. Commonly used pattern classification methods are k-Nearest Neighbor Rule (KNN) [11-15], Artificial Neural Networks (ANN) [16-19], and Support Vector Machine (Support Vector Machine, SVM) [20] etc. can be used to solve diagnostic problems. Machine learning fault diagnosis is widely used because there is no requirement for data distribution, But in the face of industrial complexity, high dimensionality of data, computing power and storage space put forward higher requirements. There are many ways to control and diagnose, For

* Corresponding author.
E-mail addresses:lvfeng@mail.tsinghua.edu.cn(F. Lv)

example, Time Delay Estimation-based Adaptive Sliding-Mode Control for Nonholonomic Mobile Robots[21], Fault Diagnosis of Analog Circuit based on Extreme Learning Machine[22], Finite Time Synergetic Control for Quadrotor UAV with Disturbance Compensation[23].

Actual industrial production is accompanied by environmental changes, material fluctuations, set point changes, instrument aging, operating mode switching, etc., resulting in the acquisition of data through multiple operating modes or working conditions, that is multi-time intermittent process. Lu et al. [24] proposed the multi-period batch process has a very valuable phenomenon: the potential variable correlation in the batch process does not change with time, but the segmentation is followed by the process operation process or the change of process mechanism characteristics. Multimodal processes with multi-center, variable non-Gaussian, nonlinear, etc. Fault detection for transition periods in multimodal processes is a more complex problem. Kosanovich et al. proposed in 1994 that production models with distinctly different correlations should establish different statistical models, so that the results will more accurately reflect the correlations and diagnostic results of different processes [25, 26]. For the transition period in multimodality, Tan et al. pointed out that its characteristics are not always in strict recursion, and there may be repetitions and exceptions sometime. The transitional modal data has many variables and complex changes. If the MPCA fault diagnosis method is used, it is necessary to establish a model for the parameter characteristics, not only does the universal diagnosis fail but also to achieve the desired effect. If the FD-KNN diagnostic method is used, the high-dimensional data characteristics increase the amount of calculation. Therefore, based on the characteristics of the transitional modal data, this paper replaces the traditional PCA dimensionality reduction with the random projection method on the basis of FD-KNN, and diagnoses the transition mode. This method has three advantages. (1) Simplify the modeling method and avoid nonlinear and non-Gaussian problems. (2) The amount of calculation is reduced by dimension reduction, and the fault diagnosis time is shortened. (3) Solving the problem of non-protection distance of PCA dimension reduction in transition mode, further mining the information of sampling data and improving the utilization of transition modal data.

2. Transitional modal data characteristics

The data in different stable states have different correlations, so it is necessary to establish different stable modal models. Different transition modes also need to establish different models. The transition mode exists from the end of the previous stable mode to the beginning of the next stable mode. The data characteristics of transition modes are different from those of steady state data. In general, the process characteristics at each time point at the beginning of the transition are close to those of the previous mode. With the continuous transition process, the characteristics gradually shift to the latter mode of operation mode[27].

Monitoring of transient mode is a difficult point in the whole process monitoring. Shuai Tan and others pointed out that[28]in the transition mode, although the operational characteristics of the data usually change in real-time, the normal transition period of the data changes have a more fixed trajectory. In other words, if the random disturbance is not considered, the variables will change regularly along a certain trend during the transition period. The relative change of variables in the same period is regular, which reveals the time-varying trend and characteristics of variables in the transition period. A trajectory with the best transition mode of AB is set as the model by statistical method. By comparing the transition mode of AB in real-time monitoring production with that of modeling, the transition mode of AB in the current time can be judged whether it is running under the normal trajectory.

Transition modes of batch processes need to add a dimension (batch) to the two-dimensional (sampling points and variables). That is, $X(I \times J \times k)$, and its three-dimensional representation: batch (transition times i, process variables j, and the number of sampling points in a batch k).

Generally speaking, the transitional data have the following characteristics: (1) There are different modes in different periods, and the modes in different transitional periods are also very different. That is, multimodality. (2) Most variables are nonlinear and inconsistent with Gauss distribution. That is, the variability of data variables. (3) The transition data are relatively small, and the correlation between different transition intervals is quite different. But the transition modes of the same sub-period have certain rules.

3. Comparison of diagnostic methods

3.1MPCA diagnostic methods

The *MPCA* method is the same as the *PCA* modeling method, which is the extension of the *PCA* method.

Firstly, the three-dimensional data is expanded by variable expansion method to form k time slices $X_k(I \times J)$, and then the data is standardized. The standardized formula is as follows:

$$X = \frac{X_k - e \cdot q}{diag(\sigma_1, \sigma_2, ..., \sigma_n)} \tag{1}$$

Where: $X_k \in \Re^{n \times d}$, $n$represents the number of samples in the kth time slice, that is the number of sampling points i, d represents the number of measured variables i n the k th time slice, that is the number of sensors j. $\boldsymbol{e}$ is a full rank matrix whose elements are all 1, $q = \{q_1, q_2, ..., q_n\}$ is the meanmatrix, $diag(\sigma_1, \sigma_2, ..., \sigma_n)$ is the covariance matrix of X_k. All of the X in the following are dimensionless standardized matrices.

The two-dimensional data X is analyzed by principal component analysis. First, the covariance is calculated. The covariance formula ofX is as follows:

$$S = \frac{1}{n-1} X^T X \tag{2}$$

Eigenvalue solution for covariance matrix，the formula for solving eigenvalues is as follows:

$$\begin{aligned} &|\lambda_i I - S| = 0 \\ &(\lambda_i I - S) p_i = 0 \end{aligned} \tag{3}$$

The eigenvalues are arranged in descending order, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n \geq 0$, eigenvalue eigenvector pairs can be obtained，（λ_1，p_1),(λ_2, p_2)...(λ_n, p_n）. Then CPV（cumulative percent variance，CPV） is used to test the number of best principal components. The total contribution rate of the preceding A principal components can be expressed as:

$$CPV = \frac{\sum_{i=1}^{A} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{4}$$

Among them, A usually refers to the number of principal elements,

and a_k is the number of principal components of the k time slice. λ is the characteristic value arranged from large to small.Usually when the principal component information in CPV is greater than 85%. The model can represent the information contained in the matrix, and then achieve the optimal dimensionality reduction.

At this point, the matrix X can decompose the two parts of the main element space and the residual space.That is:

$$X = \tilde{X} + E = TP^T + E \quad (5)$$

$$T = XP \quad (6)$$

$$\tilde{X} = TP^T \quad (7)$$

Among them, E is the residual matrix, T is the principal score score matrix, and P is the load matrix. In general, the X matrix can be decomposed into a combination of $X = t_1 p_1^T + t_2 p_2^T + \ldots t_i p_i^T$ and $i = 1, 2 \ldots n$, where $t_1, t_2 \ldots t_i$ is the score vector and $p_1, p_2 \cdots p_i$ is the load matrix. Get the load matrix $P_k (j \times a_k)$.,and $a_k < j$。

The principle of $MPCA$ modeling is shown in Figure 1. Among them, a_k is the number of principal components of the k time slice. X_k is the k time slice.

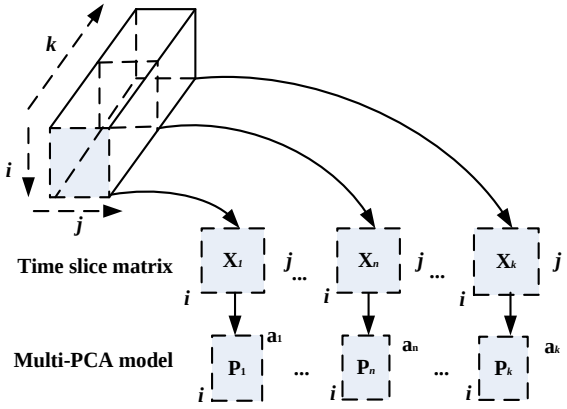


Fig. 1 Schematic diagram of MPCA modeling

Similar to the method of PCA, $MPCA$determines the state of industrial production by determining the control limits and statistics of SPE and Hotelling T^2.

$$T^2 = t_{k,i} \Lambda_k^{-1} t^T_{k,i} \quad (8)$$

Among them, $t_{k,i}$ represents the principal component variable of the i batch of the k th time slice, and the diagonal matrix Λ_k consists of the first a_k eigenvalues of the time slice k after normalization and covariance operation.

SPE Statistic is

$$SPE_{k,i} = \left(x_{k,i} - \tilde{x}_{k,i}\right)\left(x_{k,i} - \tilde{x}_{k,i}\right)^T \quad (9)$$

Among them, $E_{k,i} = x_{k,i} - \tilde{x}_{k,i}$。

T^2 Control line is

$$T^2 \sim \frac{A(I-1)}{I-A} F_{A,I-A,\alpha} \quad (10)$$

Among them, A is the number of main components, $F_{A,I-A,\alpha}$ is F distribution, α is a significant level.

SPE Control line is

$$SPE_{k,\alpha} = g_k \chi_{hk,\alpha}$$

$$g_k = \frac{v_k}{2m_k}$$

$$h_k = \frac{2m_k^2}{v_k} \quad (11)$$

Among them, m_k and v_k are the mean and variance of SPE statistics at the k sampling point.

However,in dealing with non-linear non-Gaussian problems, we need to model, analyze the data, and match the model, which is relatively inefficient.

3.2 FD-KNN method

The traditional k-nearest neighbor rule is a classical pattern classification method, which can deal with multi-modal, non-linear, non-Gaussian problems in industrial processes. And because its algorithm is simple and practical, it has been widely used in many fields[29.30]. In the field of industrial process monitoring, KNN has also been used to deal with fault classification[31.32].

He et al. retained the idea that the distance between samples was used as the difference measure in KNN classification algorithm. According to the distribution characteristics of online data distance, KNN was applied to real-time monitoring fault detection algorithm, called Fault detection method using the k-nearest neighbor rule, FD-KNN. The idea of the algorithm is that if the measured data obtained on-line is normal data, it should be very similar to the historical normal data; on the contrary, if the on-line measurement data is a fault sample, it is obviously different from the training sample. This difference is represented by the square of K nearest neighbor mean distances between training set samples and online data as monitoring statistics, i.e. the distance between nearest neighbors.

$$D_i^2 = \frac{1}{K} \sum_{i=1}^{k} d_{ik}^{\ 2} \quad (12)$$

Among them, d_{ik} is the square of the Euclidean distance between the i sample and the k nearest neighbor.

FD-KNN is able to deal with multimodal, nonlinear and non Gauss problems in industrial processes. However, FD-kNN faces the heavy computational burden of calculating the distance between high-dimensional vectors and sorting when there are more measurement variables, which reduces the efficiency of its online application.

The diagnostic process of using FD-KNN is as follows.

(1) offline modeling

a: Find K nearest neighbors in the training samples: calculate the minimum distance d_{ij} (Euclidean distance) between each sample X_i, $\mathrm{i} = 1,2,\cdots,n$ and other samples X_j, $\mathrm{j} = 1,2,\cdots,n$ in the training samples.

$$d_{ij} = \left\| X_i - X_j \right\|_{\ell_2},\ j \neq i \quad (13)$$

b: The square of the average distance between each sample and its K neighboring samples is calculated.

$$D_i^2 = \frac{1}{K} \sum_{i=1}^{k} d_{ik}^{\ 2} \quad (14)$$

c: Reference literature [16]，There are two ways to estimate thresholds. One method is to estimate D_α^2 by approximating D_i^2 to non-central chi-square distribution when d_{ik} conforms to

non-zero mean normal distribution. Another way is to divide the threshold into threshold according to the experience of D_i^2, that is the method adopted in this paper.

$$D_\alpha^2 = D^2_{\left(\lfloor n(1-\alpha)\rfloor\right)} \tag{15}$$

Among them, D_i^2, $i=1,2,\cdots,n$ is the result of descending the order of D_i^2. $\left(\lfloor n(1-\alpha)\rfloor\right)$ represents the integral part of $n(1-\alpha)$.

(2) fault diagnosis

a: Find the K neighbors of X from training samples.

b:According to (2) calculation D_x^2

c: Compare D_x^2 and D_α^2, if D, the online sample X is the fault sample; otherwise the online sample X is the normal sample.

It should be noted that FD-KNN reduces the fault diagnosis rate when calculating the distance between high-dimensional data and sorting large sample data. Although the transitional period data is relatively small, there are many transitional periods in an intermittent process, and the situation of different engineering data is different.

3.3 RP-KNN fault diagnosis method

The PCA dimension reduction method only retains the principal metaspace part and the dimensionality reduction has no distance protection. Based on this, Zhou Zhe [33] adopted the method of random projection based KNN rule (RPKNN).

Random projection is based on an important lemma, the Johnson-Lindenstrauss (JL) lemma [34]. The lemma states that for any set of points (n points) in a high-dimensional (d-dimensional) Euclidean space, the set of points can be mapped to a low-dimensional space, and the dimensionality is also ensured that the distance of the points remains approximately constant after projection while reducing the dimension, and thedistance loss ratio does not exceed ε, （$\varepsilon>0$）[35, 36]. Many experts have proposed a simpler method for JL lemma, setting the minimum bound for the low-dimensional space dimension [37.38]0Achlioptas designed two very simple projection matrices, which consist of only 0, 1 elements in this projection matrix. . This theorem makes the projection operation fast, and is described as follows:

Theorem: Q represents any set in $\Re^d$ dimensional space (this set has n points), store the set in the matrix $A\in\Re^{n\times d}$.

$$L_0 = \frac{4+2\beta}{\varepsilon^2/2-\varepsilon^3/3}\log n \tag{16}$$

R is an $d\times l$ dimensional random matrix, and the elements in the matrix are independent random variables.

The variables $\{r_{ij}\}$ in the matrix have the following two values:

$$p(r_{ij}=1)=p(r_{ij}=-1)=1/2 \tag{17}$$

$$\begin{cases} p(r_{ij}=1)=p(r_{ij}=-1)=1/6 \\ p(r_{ij}=0)=2/3 \end{cases} \tag{18}$$

Make

$$E=\frac{1}{\sqrt{L}}AR \tag{19}$$

f: $\Re^{\mathrm{d}}\to\Re^{\mathrm{L}}$ represents the mapping. the i-th row of the matrix A can be mapped to the corresponding row in the matrix E, and for any two points in the original matrix Q, u and v are at least $1-n^{-\beta}$ in the following formula.

$$(1-\varepsilon)\|u-v\|^2 \le \|f(u)-f(v)\|^2 \le (1+\varepsilon)\|u-v\|^2 \tag{20}$$

The parameter ε controls the accuracy of the distance between the point and the point after the projection, and the parameter β controls the success rate of the projection. If the ε selection is inappropriate, the training data is different from the neighbors of the online sample selection. Therefore, it is necessary to select ε [39]

Assume that the training sample set is

$$D=\begin{bmatrix} d_{1,1} & \cdots & d_{1,k} & d_{1,k+1} & \cdots & d_{1,n-1} \\ \vdots & & \vdots & \vdots & & \vdots \\ d_{i,1} & \cdots & d_{i,k} & d_{i,k+1} & \cdots & d_{i,n-1} \\ \vdots & & \vdots & \vdots & & \vdots \\ d_{n,1} & \cdots & d_{n,k} & d_{n,k+1} & \cdots & d_{n,n-1} \end{bmatrix} \tag{21}$$

Here $d_{i,j}$ represents the minimum distance $d_{i,j}$ between each sample X_i, i $=1,2,\cdots,n$ and the other samples X_j, j $=1,2,\cdots,n$.

Assume that the k-nearest neighbor indices with the number of samples i in the training set data are matrix N^{i}_{index}, and the k-nearest neighbor indices with the number of samples i in the projection matrix are the matrix $\overline{N}^{\mathrm{i}}_{index}$.

According to the following theorem:

Q represents a set of any n points in the $\Re^d$ pace, and the distance matrix D contains the Euclidean distance between any two points in the set Q. Set N^{i}_{index} contains i samples with k-nearest neighbor indexes.

According to formulas

$$p(r_{ij}=1)=p(r_{ij}=-1)=\frac{1}{2}$$

and

$$\begin{cases} p(r_{ij}=1)=p(r_{ij}=-1)=1/6 \\ p(r_{ij}=0)=2/3 \end{cases}$$

a random matrix $R\in\Re^{n\times d}$ can be generated, and the dimension ofdata points in the set Q can be reduced by the matrix R. The distance matrix of the sample set in the projection space are $\overline{D}=\left[\overline{d}_{i,j}\right]$, and the k-nearest neighbor index set of the i-th sample are $\overline{N}^{\mathrm{i}}_{index}$.

If the parameter e satisfies

$$\varepsilon \le \frac{\min\limits_{i\in\{i,\cdots,n\}}\frac{d_{i,(k+1)}}{d_{i,k}}-1}{\min\limits_{i\in\{i,\cdots,n\}}\frac{d_{i,(k+1)}}{d_{i,k}}+1} \tag{22}$$

then

$$\overline{N}^{i}_{index}=N^{i}_{index}\ i=1,\cdots,n \tag{23}$$

The KNN fault detection steps based on random projection are as follows:

Using random projection to reduce dimensionality

a: establish a random projection matrix R

b: Project the training sample set data matrix x into a random subspace: $T_{RP} = XR$ (24)

(1) Use new data T_{RP} to build a model using FD-KNN method according to Section 3.2

(2) Fault detection

a: Project the online sample data into the random subspace, assuming the online sample set is y, $t_y = R^T y$ (25)

b: find the k-nearest neighbor of t_y from the sample set T_{RP}, and calculate $\overline{D}_{t_y}^2$

c: Compare $\overline{D}_{t_y}^2$ and $\overline{D}_{\alpha}^2$ to diagnose whether the online sample data is fault data.

4. Practical application

4.1 Experimental environment

In the machining electric drive system, it mainly consists of a drive motor, a transmission mechanism, a cutting device and a control device. The cutting subsystem and the motor subsystem in the system are in an unstable condition during the cutting process of the same workpiece, and the load is constantly changing. Among them, the electric motor is an important power equipment, and the reliability of its operation is directly related to the quality of the whole system and product processing. Therefore, it is necessary to improve the accuracy of fault judgment during its intermittent operation.

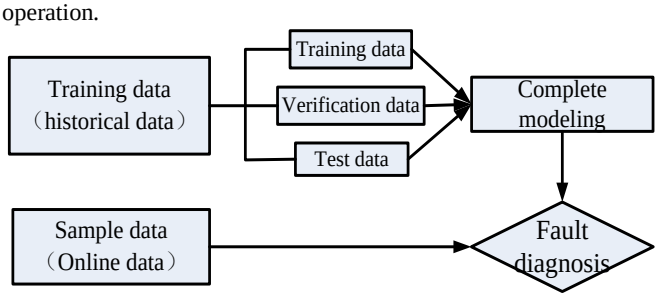


Fig. 2 data processing flow chart

Test data analysis: Real-time monitoring of the machining process of a lathe equipment. The technical data of the three-phase asynchronous motor is rated capacity of 7.5 kW, rated current of 20 amps, and rated speed of 1440 rpm. The operating parameters were measured during the fine and rough machining of the workpiece, including the three-phase current, speed, power, power factor, grid voltage, frequency, torque, temperature and other parameters of the motor stator. And it is tested under different processing materials, different feed rates, different operating conditions, etc. The data in the machining electric drag system experiment is three-dimensional data, which is the batch, which is the number of variables, and is a batch sampling point.

4.2 data processing

In order to obtain more data and shorten the sampling point time under the condition of constant production conditions, the normal transition time slice of 100 sampling points is obtained, which is divided into a training group and a verification group. 90 randomly selected from 100 time slices are used as training groups, and the remaining 10 time slices are used as verification groups. The verification group is used to verify whether the model constructed by the training group can properly process normal data. Randomly select 10 sets of time slices as test groups to join faults, and monitor the ability of the model to handle faults. In order to minimize the impact of randomness, the experimental and test groups selected for each experiment were 100 different experiments. The data processing flow is shown in Figure 2.

4.3Comparison between PCA and random projection

Using the traditional multivariate statistical method PCA diagnostic method dimension reduction method requires the establishment of covariance matrix, eigenvalue solution, computational complexity and time. The dimensionality reduction results only retain the characteristics of the wholeness. It can be understood as the geometrical retention of the direction of the larger variance of the projection, but the change of the distance between the data does not preserve the relationship between any two points [37], resulting in fewer transitions. The data is missing some of the feature information, resulting in poor diagnostic results. Based on PCA dimensionality reduction as shown in Figure 3.

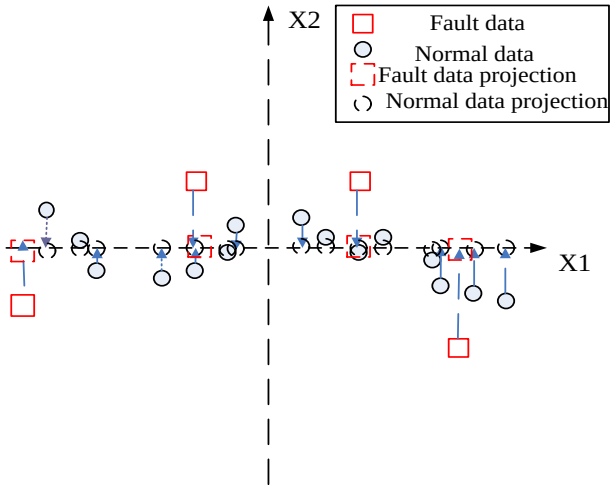


Fig.3 PCA dimension reduction diagram

Among them, the red box represents the fault data, and the grey circle represents the sample data. According to the PCA projection method, the data is projected to the X1-axis. The projected sample data is represented by dotted circles, and the fault data is represented by dotted square. It can be seen from Figure 3 that the distance information between the data has obviously changed.

Random projection method is not adaptive, it is not affected by the collected data, to achieve dimensionality reduction while ensuring the distance between samples. In order to intuitively represent the comparison of the distance preservation between PCA and random projection, the relative distance loss isused as the measure index l. Distance loss is

$$l = \frac{d_{i,j}^{p} - d_{i,j}}{d_{i,j}} \tag{26}$$

Where $d_{i,j}^{p}$ denotes the distance between i and j samples after dimensionality reduction, B denotes the distance between i and j samples. The experiment was conducted 100 times, and the distance loss test was performed in each training group. The measurement results are shown in Figure 4.

The red circle is the PCA dimension reduction distance loss value, and the blue line circle is the random projection dimension reduction distance loss value. The absolute value of PCA dimensionality reduction distance loss is between [-0.36-0.32], and the projection dimensionality reduction distance loss is between [.01-0.04]. This means that the random projection dimension reduction is relatively conservative compared with the PCA dimension reduction.

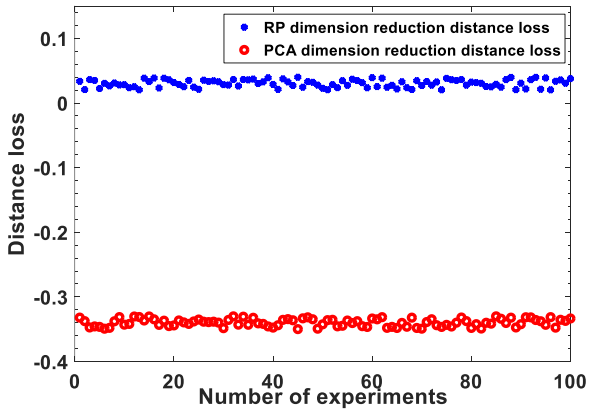


Fig.4 Distance loss comparison chart

4.4 comparison of FAR and FDR with different diagnostic methods

Sampling at intervals of one second, recording the data, that is, there are 25 time slices. Set the time slice as the fault time slice. Table 1 shows the failure false alarm rate after 100 experiments, and Table 2 shows the test results for each failure time slice at 100 times. (Note: when the result coincides with the control line, it is calculated by 0.5%). The failure rate of false alarm is defined as follows:

Definition 1: False Alarm Rate (FAR) indicates that the number of normal samples misreported as fault samples accounts for the proportion of normal samples in all verification sets.

False alarm rate = false positive number / normal sample size

Deformation2: Deformation.on Rate(Fault.on Rate,FDR)defined to the the function sample to the the function sample to the function sample to the function sample.

Fault detection rate = fault output / total number of failures

Tab.1 False Alarm Rate

Method	MPCA	FD-KNN	RP-KNN
FAR	0.045	0.05	0

Tab.2 Deformation.on Rate

Method / Fault time slice	MPCA	FD-MPCA	RP-MPCA
1	1	1	1
2	1	1	1
3	0.85	0.80	0.9
4	0.95	0.95	1
5	1	1	1
FDR	0.95	0.89	0.99

It can be seen from Table 1 that the false fault alarm rate of RP-KNN is 0, the falsefault alarm rate of MPCA is 0.045, the false fault alarm rate of FDKNN is 0.05, and the fault false alarm rate of RP-KNN is the lowest.

It can be seen from Table 2 that the false fault alarm rate of RP-KNN is 0.99, the falsefault alarm rate of MPCA is 0.95, the false fault alarm rate of FDKNN is 0.89, the fault false alarm rate of RP-KNN is also the lowest.

4.5Comparison of fault diagnosis between MPCA and RP-KNN

Four sampling points, I = 17, I = 18, I = 19, I = 77, are distributed at the early and late stages of the transition period. In these four sampling points, the fault is manually set up using MPCA and RP-KNN for fault diagnosis. Figure 5 , Figure 6is the result of MPCA fault diagnosis. Figure 7 is the result of RP-KNN fault diagnosis.

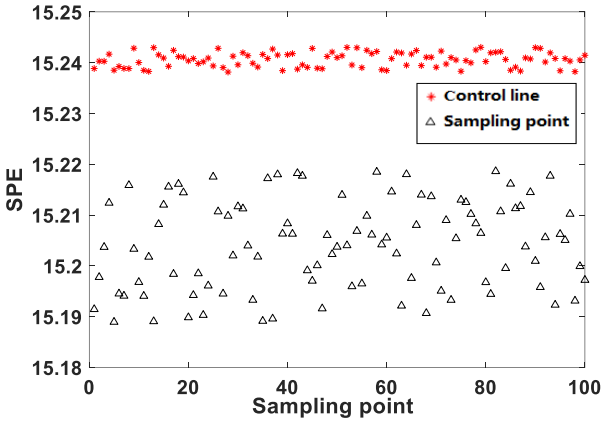


Fig.5 SPE indicators in MPCA fault diagnosis

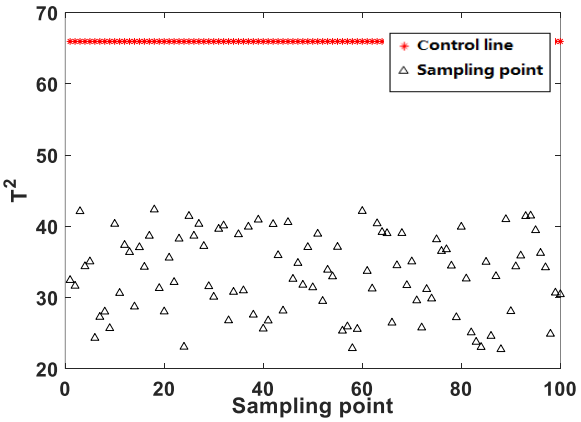


Fig.6 T^2 indicator in MPCA fault diagnosis

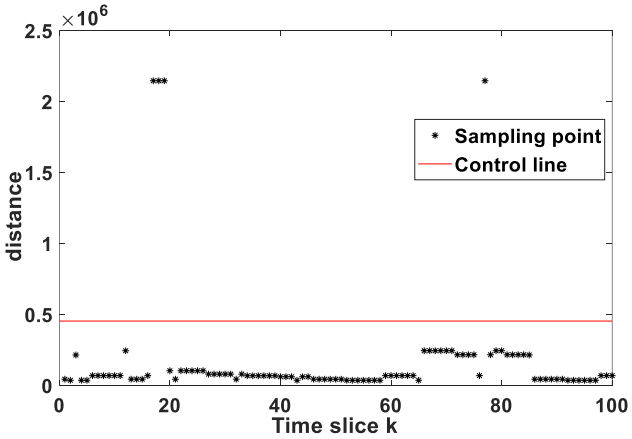


Fig.7 RP-KNN troubleshooting

Figure 5 , Figure 6 shows that a small sample of fault data exceeds the control line, and MPCA is not sensitive to fault diagnosis in transition period. Figure 7 shows that the fault data is beyond the control line. The comparison between figure 5, figure 6 and Figure 7shows that the RP-KNN diagnosis method is relatively accurate for motor fault diagnosis in transition period.

5. Summary

In this paper, a fault diagnosis method based on RP-KNN is proposed for intermittent multi-period industrial production process. The method can satisfy the transient fault detection of three-dimensional data with multi-period interval transition process, and has the following advantages:

The modeling method is simplified, which avoids the nonlinear and non Gauss problems.

It solves the problem of dimensionality reduction without distance preservation in the transitional mode by PCA, further excavates the information of the sampled data, and improves the utilization ratio of the transitional mode data.

The validity of the method is proved by the experimental simulation using the test data of machine tool processing equipment, and the research has practical application value.

Acknowledgements

The work has been partly supported by National Natural Science Foundation of China (NO.61673160, 60974063, 61175059); Natural Science Foundation of Hebei Province (F2018205102); Hebei Provincial Department of Education Project (ZD2016053); Ministry of Education “Chunhui Plan” Cooperative Scientific Research Project (Z20177023).

References

[1] S. Joe Qin. Surveyon data-driven industrial process monitoring and diagnosis. Annual Reviews in Control,2012, 36(2): 220-234.

[2] J Bhavik R. Bakshi. Multiscale pca with application to multivariate statistical process monitoring. AIChE Journal, 1998, 44(7): 1596-1610.

[3] Weihua Li, H Henry Yue, Sergio Valle-Cervantes, S Joe Qin. Recursive pca for adaptive process monitoring. Journal of Process Control, 2000, 10(5): 471-486.

[4] Xun Wang, Uwe Kruger, George W Irwin, Geoff McCullough, Neil McDowell. Nonlinear pca with the local approach for diesel engine fault detection and diagnosis. IEEE Transactions on Control Systems Technology, 2008, 16(1): 122-129.

[5] Yuan Yao, Furong Gao. A survey on multistage/multiphase statistical modeling methods for batch processes. Annual Reviews in Control, 2009, 33(2): 172-183.

[6] ZHAO Chunhui,Wang Fuli,Yao Yuan,Gao Furong.Phase-based Statistical Modeling, Online Monitoring and Quality Prediction for Batch Processes.ACTA AUTOMATICA SINICA 2010, 36(3): 366-374.

[7] ZHOU Donghua, YE Yinzhong. Modern Fault Diagnosis and Fault Tolerant Control [M] Beijing: Tsinghua University Press, 2000.

[8] HE Q P, WANG J. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes [J]. IEEE Transactions on Semiconductor Manufacturing, 2007, 20(4): 345-354.

[9] MA H H, HU Y.SHI H B. Fault detection of complex chemical processes using Mahalanobis distance-based local outlier factor [J]. CIESC Journal, 2013, 64(5): 1674-1682.

[10] Guo Jinyu,Chen Haibin,Li Yuan.KNN Fault Detection Method for Batch Process Based on Principal Sample Modeling Upgraded Online[J]. Information and control ,2014, 43(4): 495-500.

[11] CK Mechefske, J Mathew. Fault detection and diagnosis in low speed rolling element bearings part ii: The use of nearest neighbour classification. Mechanical Systems and Signal Processing, 1992, 6(4): 309-316.

[12] Roland Casimir, Emmanuel Boutleux, Guy Clerc, A Yahoui. The use of features selection and nearest neighbors rule for faults diagnostic in induction motors. Engineering Applications of Artificial Intelligence, 2006, 19(2): 169-177.

[13] Q Peter He, Jin Wang. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. IEEE Transactions on Semiconductor Manufacturing, 2007, 20(4): 345-354

[14] Yaguo Lei, Ming J. Zuo. Gear crack level identification based on weighted k nearest neighbor classification algorithm. Mechanical Systems and Signal Processing, 2009, 23(5): }535-1547

[15] Dong Wang. K-nearest neighbors based methods for identification of different gear crack Levels under different motor speeds and loads: Revisited. Mechanical Systems and Signal Processing. 2016. 70-71:201-208.

[16] Jie Zhang, PD Roberts. On-line process fault diagnosis usingneural network techniques.Transactions of the Institute of Measurement and Control, 1992,14(4}:179-188.

[17] Jie Zhang, AJ Morris. On-line process fault diagnosis using fuzzy networks. Intelli-gent systems engineering, 1994, 3(1): 37-47.

[18] Jie Zhang, Julian Morris. Process modelling and fault diagnosis using fuzzy neural networks. Fuzzy Sets and Systems, 1996, 79(1): 127--140.

[19] Jie Zhang. Improved on-line process fault diagnosis through information fusion in multiple neural networks. Computers&Chemical Engineering, 2006, 30(3): 558-571.

[20] LB Jack, AK Nandi. Support vector machines for detection and characterization of rolling element bearing faults. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2001, 215(9): 1065-1074.

[21] Xiaoyuan Luo,Shangbo Ge,Jiange Wang, Guan,Time Delay Estimation-based Adaptive Sliding-Mode Control for Nonholonomic Mobile Robots[J],International Journal of Applied Mathematics in Control Engineering,2018:1-8.

[22] Du, W,Zhao, X, Du, H, Lv, F, Fault Diagnosis of Analog Circuit based on Extreme Learning Machine [J],International Journal of Applied Mathematics in Control Engineering,2018:9-14.

[23] Qiang Chen, Yinqiang Wang, Zhongjun Hu, , Finite Time Synergetic Control for Quadrotor UAV with Disturbance Compensation [J],International Journal of Applied Mathematics in Control Engineering,2018:31-38.

[24] Lu N Y, Gao F R, Wang F L. A sub-PCA modeling and on-line monitoring strategy forbatch processes. AIChE J, 2004, 50(1):255-259.

[25] Kosanovich K A, Piovoso M J, Dahl K S, MacGregor J F, Nomikos P. Multiway PCA applied to an industrial batchprocess. In: Proceedings of the American Control Conference. Washington D. C., USA: IEEE, 1994. 1294-1298

[26] Kosanovich K A, Dahl K S, Piovoso M J. Improved processunderstanding using multiway principal component analysis. Industrial and Engineering Chemistry Research, 1996, 35(1): 138-146

[27] Bhagwat A, Srinivasan R, Krishnaswamy P R. Multi-linearmodel-based fault detection during process transitions. Chemical Engineering Science, 2003, 58(9): 1649-1670

[28] Tan Shuai, Wang Fu, Chang Yuqing,Wang Shu,Zhou He. Fault Detection of Multi-mode Process Using Segmented PCA Based on Differential Transform [J]. ACTA AUTOMATICA SINICA ,2010,36(11),1627-1635

[29] Huang Jie, Zhu Guangping. Research on the K-D Tree KNN 一 SVM Classifier in Underwater Acoustic Target Recognition[J].Acta Oceanologica Sinica,2018,37(01):15-22.

[30] Huang Xianying, Xiong Liyuan, Liu Yingtao, Li Yidong. An improved KNN short text classification algorithm based on category feature words [J].Computer Engineering & Science,2018,40(01):148-154.

[31] Chen Fafa, Li Wei, Chen Baojia, Chen Congping..Fault Diagnosis of R oller Bearing based on Hybrid Feature Set and Weighted KNN [J]. Mechanical transmission,2016,40(08):138-143.

[32] Zhang Cheng, Guo Qingxiu, Li Yuan..Varimax rotationbased K nearest neighbors fault diagnosis strategy [J/OL].Application Research of Computers,2019(09):1-10

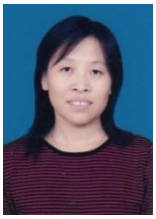
[33] Zhou Zhe. Fault diagnosis of complex industrial processes based on k-nearest neighbors [DZhejiang University PhD thesis.2016

[34] William B Johnson, Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. Conf. in Modern Analysis and Probability, volume 26. 1984:189-206.

[35] DimitrisAchlioptas. Database-friendly random projedtions.Proc.ofthe twentieth ACMSIGMOD-SIGACT-SIGART symposium on Principles ofdatabase systems.Montreal,Quebec,Canada,May.2001;274-281
[36] D Sivakumar. Algorithmic derandomization via complexity theory. Proceedings of the thiryfourth annual ACM symposium on Theory of computing. Montreal, Quebec, Canada, May 2002:619-626.
[37] Peter Frankl, Hiroshi Maehara. The johnson-lindenstrauss lemma and the sphericity of some graphs. Journal of Combinatorial Theory, Series B, 1988, 44(3): 355-362.
[38] Sanjoy Dasgupta, Anupam Gupta. An elementary proof of a theorem of Johnson and linden-strauss. Random Structures&ALgorithms, 2003, 22(1):60-65
[39] Yang Jing, driving, Zhang Jianpei. Dimension reduction for data streams based on Johnson-Lindenstrauss transform [J]. Journal of Jilin University.2013,43(6)

Shao Mengya hebei Normal University, graduate student of physical science and information engineering. Her main research interests are in area of industrial intelligent control, information processing, fault detection and diagnosis technology and so on.

Lv Feng obtained her bachelor's degree from electrical automation major of Hebei Electromechanical College in 1982. She obtained her Master's degree from test measurement technology and instrument major of YanShan University in 2006. She is currently working in Hebei Normal University as professor and master supervisor. Her main research interests are in area of industrial intelligent control, information processing, fault detection and diagnosis technology and so on.



Du Wenxia obtained her master degree and doctor degree of control theory and control engineering major from the institute of electrical engineering of YanShan University in 2001 and 2012 respectively. She is currently working in Hebei Normal University. Her main research interests are in area of neural network control and predictive control of nonlinear system, the fault diagnosis based on data driven is her new research direction in recent years.

Song Xuejun, Master of Radio Physics and Radio Electronics, Xi'an University of Electronic Science and Technology, 1989 Received a Ph.D. in Weapon Systems and Applications Engineering from the 2009Ordnance Engineering College. He is currently a professor at Hebei Normal University and a master's tutor. The main research direction: automatic control, intelligent monitoring and fault diagnosis technology.