Research on Sensor Fault Diagnosis Method Based on KPCA-AE Algorithm

Jiongwei Fan [1,2][0009-0008-2443-6534] and Jia Zhang [1,2][0000-0002-7792-2308]

1 School of Automation, Beijing Institute of Technology, Beijing, China
2 National Key Laboratory of Autonomous Intelligent Unmanned Systems, Beijing Institute of Technology, Beijing 100081, China
zhangjia@bit.edu.cn

Abstract. Sensors occupy a pivotal role in indoor fire detection, and any malfunction within their operation can potentially escalate into severe accidents. Multi-sensor networks utilized in indoor fire early warning systems are inherently characterized by their high dimensionality and intricate linear interdependencies, presenting substantial operational complexities. To address these complexities, the paper introduces a fault diagnosis method based on Kernel Principal Component Analysis combined with Autoencoders (KPCA-AE). This methodology harmoniously integrates KPCA and AE, where KPCA is utilized to reduce the dimensionality of the data, thereby facilitating the training of the AE model. The proposed method adeptly and precisely identifies faults based on reconstruction errors, introducing a novel and effective paradigm for fault diagnosis in indoor fire sensors.

Keywords: Fault diagnosis; Sensors; Autoencoder; Kernel principal component analysis

1 Introduction

Due to the small indoor space and dense population, indoor fires can easily cause significant losses. Therefore, the prevention of indoor fire is crucial. Indoor fire warning relies on sensors. If there is fault in sensors, it will lead to data acquisition errors and monitoring failure, resulting in serious consequences [1]. Consequently, research on fault diagnosis method of indoor fire sensor has an important practical significance.

Common sensor fault diagnosis methods are mainly divided into model-based, knowledge-based and data-driven methods [2]. Due to the complexity of sensor characteristics and mechanism model, it is difficult to deal with complex sensor fault sorely through modeling or empirical knowledge. Sensors can record abundant data, which can reflect the state of the sensor. Therefore, using data-driven method to extract the hidden information of data has become the primary means to diagnose sensor fault.

Among data-driven fault diagnosis methods, the method based on machine learning is the most widely used, focusing on support vector machines and neural networks.

Support vector machine (SVM) is a classifier with good generalization ability [3]. Its performance is greatly affected by parameters, so appropriate parameters should be used in training. Aiming at the defect of many fixed parameters of traditional SVM, Qi [4] proposed a SVM fault diagnosis method based on snake optimization algorithm. It

optimized fixed parameters within a range of values and avoided getting stuck in local extrema, thereby improving the classification accuracy of SVM. To address the issue of parameter selection and validation, Li N *et al.* [5] proposed a signal recognition method based on multi-scale wavelet transform, using an improved particle swarm optimization algorithm for parameter optimization. By optimizing the penalty parameters and hyperparameters, the method can achieve higher recognition accuracy. However, the advantage of SVM method is to deal with small samples. When faced with massive data, the training time of SVM increases significantly, resulting in low efficiency.

Neural network is more suitable for processing large sample data. In recent years, researchers tend to combine neural networks with other algorithms, explore fault diagnosis methods with better comprehensive performance. Ma *et al.* [6] designed a sensor data calibration module in convolutional networks, introducing jump connections and auxiliary loss functions to calculate the fault classification and fault parameters simultaneously. Huan W *et al.* [7] proposed a residual pulse neural network that optimizes gradient transfer efficiency to achieve deep-level pulse information encoding, introduced a membrane learnable mechanism. This network demonstrates excellent information processing potential. Neural network has outstanding performance in fault classification, with good real-time performance, but it usually requires a large amount of data for training. For the problem of insufficient training caused by small samples, Sun [8] used deep convolutional generative adversarial networks to learn from imbalanced samples, effectively expanding the small dataset. Compared to SVM, neural network is able to dynamically learn and update parameters, resulting in stronger robustness.

In general, data-driven method is highly consistent and effective for sensors with rich data. This method can excavate hidden features in data and identify faults, which has become an important research direction in the field of fault diagnosis.

Aiming at problems of high data dimension, complex data correlation, and difficulty to establish mathematical model of indoor fire sensor, this paper proposes an indoor fire sensor fault diagnosis method based on KPCA-AE algorithm. This method aims to effectively reduce the data dimension, improve the efficiency of model training, and accurately diagnose sensor faults.

2 Principle of Fault Diagnosis Model Based on KPCA-AE

Autoencoder (AE) [9] is a commonly used shallow neural network in data-driven sensor fault diagnosis methods. It is an unsupervised learning algorithm that has outstanding advantages in processing unlabeled datasets. Therefore, AE is selected as the training algorithm for the fault diagnosis model. The essence of the encoder in AE is the principal component analysis method. Previous studies usually used AE as a feature extraction tool, and then combined with other classification algorithms to diagnose faults. However, for the high-dimensional dataset in this paper, it is complex to directly use AE for feature learning, which requires a long training time and is prone to over fitting. Since kernel principal component analysis (KPCA) excels in nonlinear feature extraction, whereas AE is adept at feature learning and data reconstruction, this paper

incorporates KPCA as a preprocessing step for AE. This significantly enhances the ability to reduce data dimensionality and improves the efficiency of AE training.

The workflow of KPCA-AE model fault diagnosis is shown in **Fig. 1**. The input of the model is the high-dimensional time series data of the sensor network. Firstly, the input data is filtered and standardized for preprocessing. Then, KPCA is applied to extract features and reduce the high-dimensional data to 4 dimensions. The features are then reconstructed by AE model, comparing the reconstruction error with the fault threshold to quickly determine whether the sensor is faulty and annotate the faulty data.

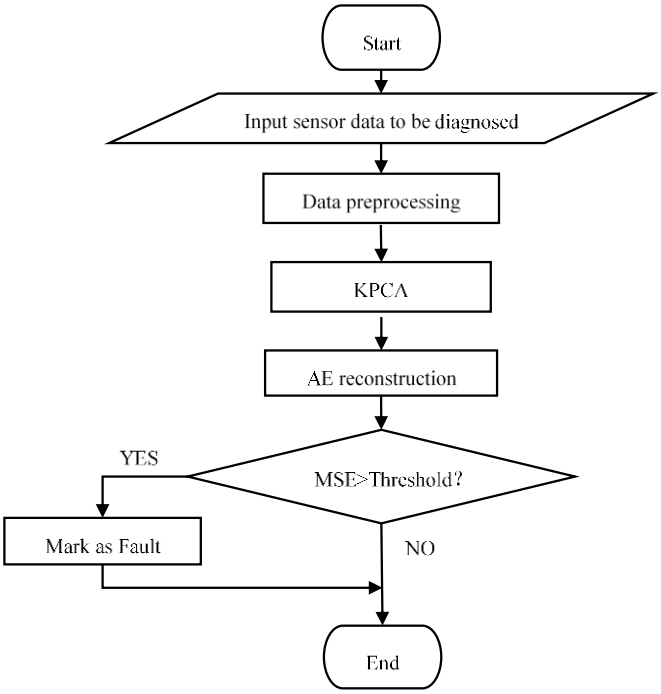


Fig. 1. Workflow diagram of fault diagnosis model

2.1 Data Source

The dataset used in this paper is from the National Institute of Standards and Technology (NIST) Building and Fire Research Laboratory's experiments on different fire alarms responding to indoor fires [10]. This experiment recorded timing data of all sensors during the fire extinguishing process. The final results of 24 experiments were saved in 24 datasets, totaling 59,421 sets of temporal data.

Each dataset contains sensors data of a fire extinguishing experiment. Sensors are distributed throughout the house. At the same time, due to the need of repeated tests, sensors of the same kind are set at the same position to reduce the error. Therefore, each experimental dataset contains over 110 dimensions, making it a high-dimensional dataset.

In addition, there are complex linear correlations between data, which will produce redundant information. Directly training will increase the complexity of the model and also reduce its accuracy. Therefore, it is necessary to extract features of original data, remove the redundant information, and reduce data dimensions.

2.2 Data Preprocessing

In the actual environment, external interference can generate noise data. Therefore, it is necessary to filter the data. In addition, the dimensions of data recorded by different sensors are inconsistent, and the values of various sensors differ greatly. As a result, it is also essential to standardize the data and unify the dimensions. To ensure the reliability of the results, SG filter is used to eliminate the noise of the original data, followed by the Z-score method to standardize the data. After data preprocessing, a clean and standardized dataset will be obtained.

2.3 Data Dimensionality Reduction

The sensor dataset contains a large number of features. It is necessary to reduce the data dimension and retain only the main features of the data, which is convenient for subsequent learning and training. KPCA is a nonlinear feature extraction method. The process of the KPCA algorithm is as follows:

KPCA dimension reduction algorithm steps.

Step 1. Complete the preprocessing of the original data. Obtain the processed dataset $\hat{X}_{i,m\times n} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n), i = 1, 2, \ldots, 24$. Where, i is the number of the dataset, and the i^{th} dataset has m time series data and n dimensions.

Step 2. Introduce the kernel function K, which is an important parameter of KPCA algorithm, and can map the original nonlinear dataset to a higher dimensional space. In this paper, the Gaussian kernel function is selected, and the function form is as follows.

$$K(x, y) = \exp\left(-\frac{\|x - y\|}{2\sigma^2}\right) \tag{1}$$

Where x and y are two arbitrary samples in the dataset, which are used to calculate the spatial distance between them, and σ is the width parameter of the Gaussian kernel function. Calculate the kernel matrix $\boldsymbol{K}$, centralize it so that the average value of each row and column of the matrix is 0, then obtain the centralized kernel matrix $\boldsymbol{K_c}$.

Step 3. Calculate the eigenvalues λ_j and corresponding eigenvectors u_j of the $\boldsymbol{K_c}$.

Step 4. Determine the principal components. The first k（k<n）principal components are selected according to the eigenvalues, which correspond to the most important features in the data. The principal components are determined by the cumulative contribution rate $G(m)$ of eigenvalues.

$$G(m)=\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{n}\lambda_n} \tag{2}$$

When taking $G(m)\geq 0.7$, the first k eigenvalues $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_k$ are selected so that the cumulative contribution rate is more than 70%, and the corresponding k principal components are determined to obtain the orthogonal basis, represented as $U=(u_1, u_2,\ldots u_k)$. U is a low dimensional space with dimension k, which is composed of the first k eigenvectors with the largest contribution of the $\boldsymbol{K}_c$ matrix.

Considering the proportion of contribution, takes k=4 to make $G(m)>70\%$.

Step 5. Construct the projection matrix, convert the data into the 4-dimensional new space, and obtain the principal components, as shown in the formula (3).

$$Y=\boldsymbol{K}_c\cdot U \tag{3}$$

Among them, as a 4-dimensional data, Y matrix is the dimension reduction result, and 4-dimensional time series data are the four principal components of the original data.

Dimension Reduction Results and Analysis.

After the dimension reduction, a low-dimensional representation of the high-dimensional dataset is obtained. At the same time, the low-dimensional representation is highly abstract, and the actual physical meaning disappears. Therefore, the principal components are numbered from high to low according to their proportion of contribution.

Taking the dataset named sdc06 as an example, the sdc06 dataset contains 117 dimensions, which are reduced to 4 dimensions using the KPCA method. The four principal component time series data curves are shown in **Fig. 2**.

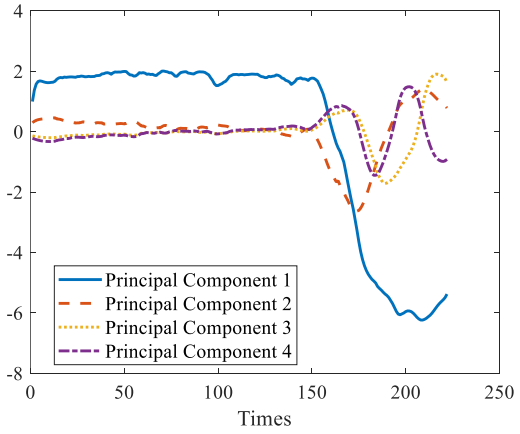


Fig. 2. Four principal component time series data curves

Draw the pie chart of the contribution of each component in the sdc06 dataset after dimension reduction, as shown in **Fig. 3**. Components are ranked from large to small

according to their contribution, and the four components with the largest contributions are selected as the dimensionality reduction results.

It can be seen that the sum of the four principal components accounts for more than 70% of all components, while the remaining 113 dimensions account for 25% in total, with their respective proportions less than 1%. The remaining components can be ignored as minor. Through KPCA dimensionality reduction, the high-dimensional original data is reduced to four dimensions. Using the 4 principal components to describe the dataset can not only retain the main characteristics of the dataset, but also reduce the difficulty of subsequent model training.

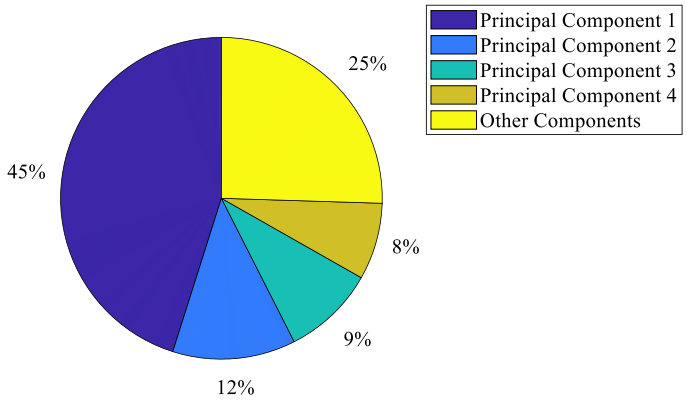


Fig. 3. The proportion of contribution of components

3 Training of Sensor Fault Diagnosis Model Based on AE

3.1 Dataset Partitioning

In order to avoid the problem of over fitting, sufficient experiments are needed. The common machine learning model evaluation method: *k*-fold cross validation method [11] is selected to divide the dataset into training and validation sets.

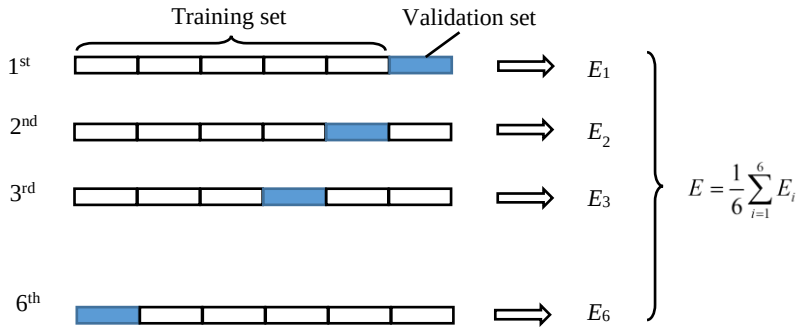


Fig. 4. 6-fold cross validation division diagram

The *k* value is generally taken as 3-10. In order to facilitate the division of 24 groups of sample data, let *k*=6 in this paper. On the original dataset, 24 groups of sample data are randomly divided evenly six times, one of which is used as the validation set and the other five as the training set. A round of training is conducted for each randomly divided training set, and the training results after 6 times of division are obtained. Generally, the average of the evaluation results is taken to obtain the final reconstruction error. The 6-fold cross validation dataset division diagram is as shown in **Fig. 4**.

3.2 AE Training Process

AE training is an iterative process, and its essence is the training of neural network. It is necessary to correctly organize the data, set up the appropriate loss function and optimizer [12]. The training steps are as follows:

(1) Data preparation and standardization.

After the KPCA data dimensionality reduction method, the dimension of the original data is reduced to four dimensions. The training set and the validation set are divided by the 6-fold cross validation method, and the training set is represented by *Y*, $Y=\left(y_1, y_2, y_3, y_4\right)$. That is, the input layer has four nodes.

(2) AE neural network creation.

Set the parameters of AE neural network and build autoencoder. The encoder and decoder of AE use sigmoid activation function, as shown in formula (4). Input the training set data samples to the encoder, the encoder maps the data to the hidden layer. The hidden layer node is set to 7, that is, the encoder can learn 7-dimensional features.

$$f\left(x\right)=\frac{1}{1+e^{-x}} \tag{4}$$

The encoded representation of the hidden layer is sent to the decoder, and the decoder restores it, represented by $\overline{Y}=\left(\overline{y}_1, \overline{y}_2, \overline{y}_3, \overline{y}_4\right)$. $\overline{Y}$ is the reconstruction result of *Y* by AE. They have the same data structure, so the output layer is also 4 nodes.

(3) Loss calculation.

The model loss between the reconstructed data and the original data is calculated. Mean square error (MSE) is used as the loss function.

$$\mathrm{MSE}=\frac{1}{m}\sum_{i=1}^{m}\left(Y-\overline{Y}\right)^2 \tag{5}$$

Where *m* is the amount of data in one dimension of the data sample. The data whose reconstruction error exceeds the error threshold is marked as fault data.

(4) Back propagation.

Back propagation algorithm is applied to adjust the weights and parameters of encoder and decoder to minimize the loss.

(5) Cycle training.

According to the training process, set the maximum number of iterations *n*. The number of iterations per round of training is about 2000. Setting the maximum number of iterations *n*=4000 can obtain appropriate reconstruction errors and avoid overfitting

or underfitting problems. Repeat steps above until the indicators of AE or reaches the limit of the maximum number of iterations *n*.

After training, the performance of AE will be verified by the validation set. The final indicator requirement is to achieve the accuracy of fault diagnosis model more than 95%, and the proportion of fault data in normal dataset is less than 5%.

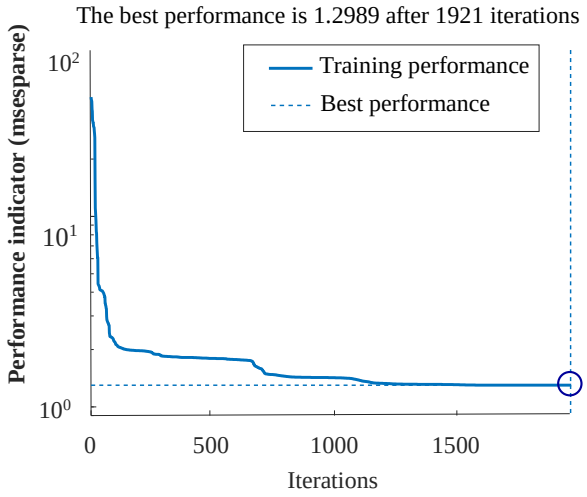


Fig. 5. AE model performance

The effect of one round of AE model training is shown in **Fig. 5**. It shows that the number of iterations is 1,921, and the training of the AE model is effective. The performance reaches its optimal value within the maximum number of training epochs, and there is no sign of overfitting or underfitting, so the dataset is fully utilized to complete the training.

3.3 Fault Diagnosis Threshold Setting Based on MSE

In the training process, it is necessary to optimize the AE structure to achieve ideal experimental results. An important evaluation index of AE model performance is MSE. MSE can be used to set a fault threshold and judge whether the input is abnormal. The process of determining the threshold is as follows:

(1) Using the empirical formula: threshold = average reconstruction error + $k \times$ standard deviation of reconstruction error. The average reconstruction error refers to the average value of the reconstruction error of all samples; Standard deviation reconstruction error refers to the standard deviation of reconstruction error of all samples; k is a parameter to control the tightness of the threshold. The larger the value of k, the looser the threshold, and only obvious abnormal data can be screened out; The smaller the value of k, the stricter the threshold, and more abnormal data can be filtered out.

(2) Based on the experimental and indicator requirements, the threshold is adjusted. The normal dataset is used as input during the training process, and the error rate needs

to be controlled within 5%. The error threshold is obtained by adjusting the coefficient according to the reconstruction results of the validation set.

The error threshold parameters are as shown in **Table 1**.

Table 1. Error threshold parameters

Average reconstruction error	Standard deviation reconstruction error	Threshold
0.3498	0.0489	0.7410

4 Simulation Experiment of Indoor Fire Sensor Fault Diagnosis

4.1 Effect of Sensor Failure on Data

The influence of sensor faults is mainly reflected in failure fault, offset fault, gain fault, impact fault, etc. [13] Sensor failure causes the detection to be missing, and the output data usually appears as 0 or NaN. Offset fault causes a certain numerical offset on the basis of the original data. Gain fault has a gain effect on the original data. Impact fault is the peak of time series data curve.

The four types of simulation faults are randomly combined into the normal dataset, and then the fault diagnosis experiment is carried out.

4.2 Experimental Results and Indicator Analysis

The fault diagnosis results of the normal dataset named sdc07 are as shown in **Fig. 6** and **Fig. 7**.

Fig. 6 is the reconstruction error of each point of sdc07 after fault model diagnosis. **Fig. 7** is the time series curve of the four principal components. The location of the fault is marked with a hollow circle. The dotted line represents the error threshold, and the data points with reconstruction error higher than the threshold are marked as fault data. Among the 570 time series data, the number of fault points was 17, accounting for 2.98%, and within 5%, which is a normal dataset.

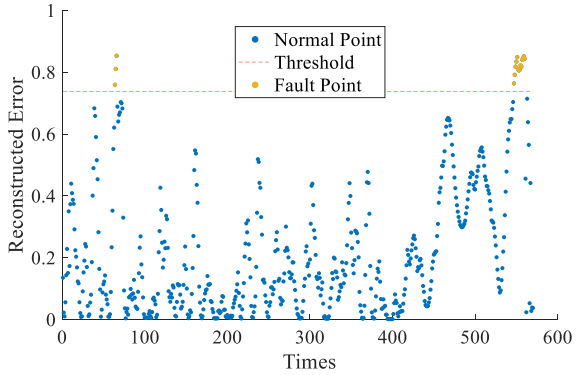
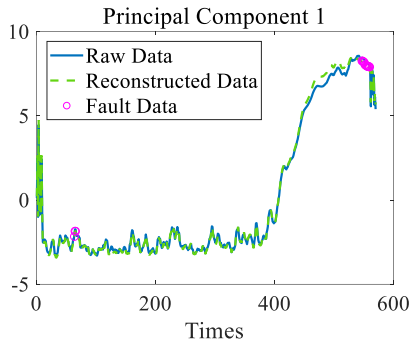
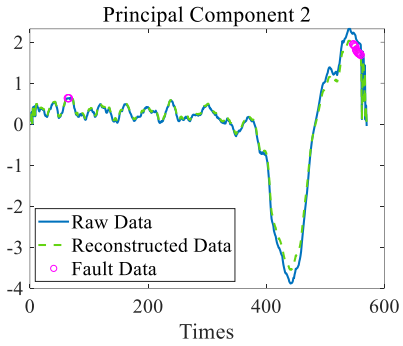


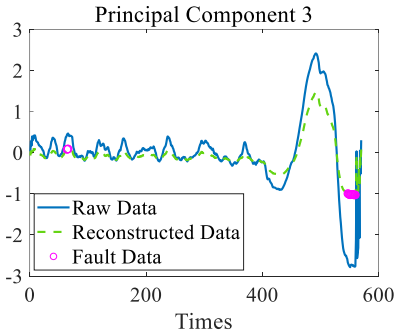
Fig. 6. Fault diagnosis reconstruction error results for normal datasets



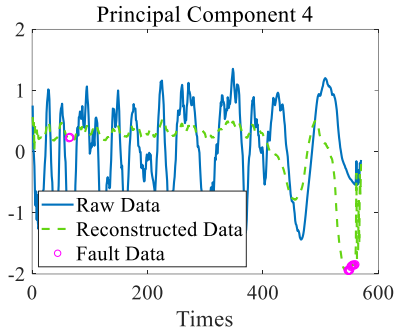
(a) Time series data of principal component 1



(b) Time series data of principal component 2

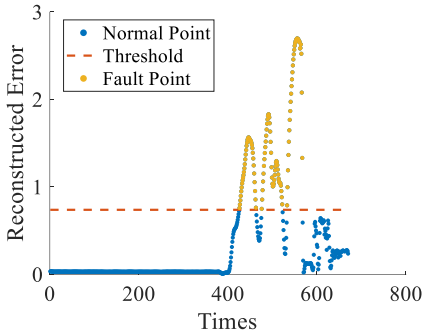


(c) Time series data of principal component 3

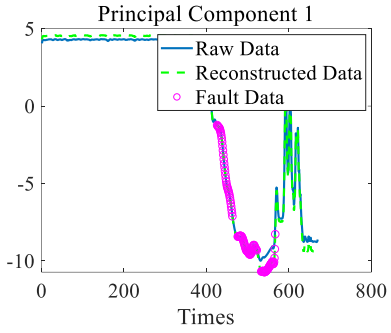


(d) Time series data of principal component 4

Fig. 7. Time series data of four principal component for fault diagnosis in the normal dataset



(a) Reconstructing error results



(b) Time series data of principal component 1

Fig. 8. Result of first fault diagnosis after adding 100 fault points

100 fault points were added to the end of the normal dataset sdc07. The result of the fault diagnosis simulation experiments is shown in **Fig. 8**. A total of 98 new ones were

added, that is, the accuracy of fault diagnosis was 98%. The subsequent experimental results are recorded in **Table 2**.

Table 2. Add 100 fault point sensor fault diagnosis results

No.	Number of fault points	Number of new faults	Fault diagnosis accuracy	Detection time/s
1	115	98	98%	5
2	104	87	87%	6
3	117	100	100%	6
4	116	99	99%	6
5	112	95	95%	5
6	117	100	100%	7
7	111	94	94%	6
8	116	99	99%	6
9	116	99	99%	5
10	117	100	100%	6
Average	114.1	97.1	**97.1%**	**5.8**

500 fault data were added to another larger normal dataset sdc11, which has 2,625 sampled data and 93 fault points. The rate of the abnormal data is 3.54%. The experimental results of fault diagnosis are recorded in **Table 3**.

Table 3. Add 500 fault point sensor fault diagnosis results

No.	Number of fault points	Number of new faults	Fault diagnosis accuracy	Detection time/s
1	583	490	98.0%	70
2	593	500	100.0%	77
3	549	456	91.2%	69
4	593	500	100.0%	89
5	563	470	94.0%	94
6	592	499	99.8%	81
7	563	470	94.0%	83
8	587	494	98.8%	83
9	568	475	95.0%	113
10	593	500	100.0%	112
Average	578.4	485.4	**97.1%**	**87.1**

The average accuracy of fault diagnosis of the two groups of experiments was more than 95%, and the standard deviation of the accuracy was 4.1% and 3.2% respectively. In the experiment, the average fault diagnosis time for each data was 0.01s. The fault diagnosis time is far less than the sampling period, which ensures that the fault diagnosis can be completed quickly in a very short time after the data is collected.

5 Conclusion

Addressing the challenges posed by high-dimensionality and intricate data correlations within indoor fire sensor data, this paper presents a sensor fault diagnosis model based on the KPCA-AE algorithm. This methodology adeptly extracts salient features while significantly reducing data complexity. Consequently, it streamlines the training process of the fault diagnosis model, alleviates computational demands, and simplifies model intricacy, ultimately yielding precise and reliable fault diagnosis outcomes. Notably, the KPCA-AE method demonstrates a fault diagnosis accuracy exceeding 95% when confronted with complex sensor data, enabling it to efficiently and promptly conduct fault diagnosis for extensive sensor datasets with stability.

References

1. Yin Z, Hou J.: Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. Neurocomputing 174(22), 643-650(2016)
2. Wu H, Zhao J.: Deep convolutional neural network model based chemical process fault diagnosis. Computers and Chemical Engineering 115(04), 185-197(2018)
3. Cortes C, Vapnik V.: Support-Vector Networks. Machine Learning 20(3):273-297(1995)
4. Qi C. An open-circuit fault diagnosis method of inverter IGBT based on data-driven. Dalian JiaoTong University (2023)
5. Li N., Zhu L., Ma W., et al.: The Identification of ECG Signals Using WT-UKF and IPSO-SVM. Sensors 22(05), 1962(2022)
6. Ma L., Guo J., Wang S., et al.: Multi-Source Sensor Fault Diagnosis Method Based on Improved CNN-GRU Network. Transaction of Beijing Institute of Technology 41(12), 1245-1252(2021)
7. Huan W., Yan Fu L.: Bioinspired membrane learnable spiking neural network for autonomous vehicle sensors fault diagnosis under open environments. Reliability Engineering and System Safety 233, 109102(2023)
8. Sun Y. Research on Fault Diagnosis Methods of Semiconductor Hydrogen Sensor. Heilongjiang University (2022)
9. Rumelhart D E., Hinton G E., Williams R J.: Learning representations by back-propagating errors. Nature 323(6088), 533-536(1986)
10. Bukowski R., Peacock R., Averill J D. et al.: Performance of Home Smoke Alarms, Analysis of the Response of Several Available Technologies in Residential Fire Settings. Technical Note https://www.nist.gov/el/nist-report-test-fr-4016, last accessed 2024/7/12.
11. Stone M.: Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society, Series B(Methodological) 36(2), 119-132(1974）
12. Zhang Y.: Research on Machinery Fault Diagnosis Method based on Deep Auto-encoder. Huazhong University of Science and Technology (2020)
13. Wang L.: Research on Fault Diagnosis Method of Flight Control System Based on Data Drive. Harbin Engineering University (2024)