IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE RELIABILITY SOCIETY SECTION

Received November 30, 2020, accepted December 15, 2020, date of publication December 18, 2020,
date of current version January 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3045815

Research on a Pattern Recognition Method of Cyclic GMM-FCM Based on Joint Time-Domain Features

YANFENG LI[1], ZHIJIAN WANG[1,2], TIANSHENG ZHAO[3], AND SONG WANQING[4]

[1]School of Mechanical Engineering, North University of China, Taiyuan 030051, China
[2]School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[3]The 713 Research Institute of CSIC, Zhengzhou 450015, China
[4]School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Corresponding author: Zhijian Wang (wangzhijian1013@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51905496.

ABSTRACT The safety and reliability of the mechanical system in the industrial process determines the quality of products. Whether the fault can be identified and classified in time is the key to ensure the safe operation of the system and arrange the appropriate maintenance plan to restrain the deterioration of the fault. However, with the rapid development of manufacturing digitization, how to process large amounts of data quickly and accurately is faced with many problems. In this paper, a pattern recognition method of cyclic GMM-FCM (CGF) based on joint time-domain features is proposed. Firstly, the concept of joint time-domain features based on Vold-Kalman filter (VKF) is proposed. It retains the integrity of the signal components and avoids the problem of dimension disaster caused by anomaly detection, which laid a foundation for the accurate classification of sensitive feature sets. Secondly, a pattern recognition method of cyclic GMM-FCM is proposed. It can eliminate global and local outliers in sensitive feature sets and determine the number of FCM categories adaptively. It makes the classification result more reasonable and accurate. Finally, the effectiveness and superiority of the pattern recognition algorithm are verified by the gearbox vibration experiments in various states. The result shows that the method is feasible in engineering practice.

INDEX TERMS Joint time-domain features, anomaly detection, GMM, FCM, pattern recognition.

I. INTRODUCTION

In the modern industrial system, with the rapid development of manufacturing digitization, real-time recording and perception of production operation state and operating environment have been realized, and a large number of industrial timing data have been accumulated and are being generated [1], [2]. In the face of massive data, how to quickly find sensitive feature sets and accurately identify and classify them is the key to efficiently discover and prevent mechanical system faults and avoid serious damage [3]. It is also the key to the research object of fault diagnosis [4]. The research on the characteristics of industrial data shows that due to the abnormal problems in the manufacturing system [5] (such as product quality defects, equipment failures, performance degradation and changes in external environment [6]), in the analysis of a large number of data, there are abnormal data which are far away from other observed data and may have different production mechanisms. The existence of such abnormal data will have a great influence on the selection and pattern recognition of sensitive feature sets [7].

For scenarios of large rotating machinery data, label of data are difficult and expensive to obtain [8]. And the marking process of label data is very dependent on subjective judgment of human. It will have great deviation and influence on the final analysis result. Unsupervised anomaly detection only relies on data without tags during training [9]. It can make use of the overall characteristics of the data to get accurate rules for dividing anomaly. Compared with the machine learning method based on label, unsupervised anomaly detection algorithm has a better application prospect [10].

However, unsupervised anomaly detection is faced with the challenge of "dimensional disaster" [11] in processing high-dimensional data, so it is necessary to select

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Li.

low-dimensional sensitive feature sets from a large number of high-dimensional data features firstly. Common processing methods include principal component analysis (PCA), feature mapping, competitive learning, topological mapping and tensor analysis [12]. However, when these methods convert high-dimensional data to low-dimensional data, some information of the signal is often lost, which affects the global outlier or local outlier removal of feature sets [13], [14]. In this paper, the single harmonic component signal obtained by Vold-Kalman filter (VKF) is used to select multiple combined features and independent features for experiment. The results show that the single combined feature sets or independent feature sets cannot be used for effective pattern recognition. However, there is still the problem of outliers in pattern recognition by topological mapping. In this paper, the signal separation results of Vold-Kalman filter are used to construct a two-dimensional joint time-domain feature sets by taking the reconstructed harmonic signal and signal residue as dimensions and taking the same time-domain features of each dimension as values. The selected sensitive feature sets not only retain the complete information of signal components, but also reduce the feature dimension, which lays a great foundation for the anomaly detection of feature sets.

For data sets without anomaly, the application purpose of machine learning in fault diagnosis is to classify the feature set and to determine whether and what kind of fault occurs [15]. The mining of data information without label is called unsupervised learning. Cluster analysis is the main method of unsupervised learning [16]. Compared with other pattern recognition methods, clustering analysis has a unique advantage in the big data scene, that is, without any known category label, the clustering analysis algorithm realizes the correct classification of samples according to sample similarity or probability density function estimation method by analyzing the internal structure of sample data. At the same time, it can be used as an independent tool to obtain the data distribution state, and can also observe the characteristics of each cluster and analyze the required cluster. Cluster analysis can be divided into hard cluster and soft cluster. Hard clustering is a well-defined clustering, that is, each sample data clearly belongs to a certain category. Soft clustering, known as fuzzy clustering, is based on the membership degree of sample data in various categories, and the category boundary is fuzzy [17]. It provides a good platform for pattern recognition of composite faults and early faults of rotating machinery.

Fuzzy C-means (FCM) [18] algorithm is the most famous and widely used fuzzy clustering method. It is very sensitive to parameters and the number of categories determines how close the clustering result is to the real data structure. If the number of classes is larger than the true value, one or more good compact clusters may be broken. If less than the true value, the clustering result merges multiple classes [17]. Therefore, how to determine the number of categories of FCM is the key to the good application of FCM.

To solve the above problems, this paper proposes a pattern recognition method of cyclic GMM-FCM based on joint time-domain features, where, cyclic GMM-FCM is named CGF. The main contributions are as follows:

(1) A method of joint time-domain features based on the VKF is proposed as the feature sets of pattern recognition. This method not only ensures the integrity of the signal component information contained in the feature sets, but also effectively reduces the dimension of the feature set and avoids the dimension disaster when the feature set is abnormal detected.

(2) The abnormal detection method of cyclic GMM-EM (Gaussian mixture model, Expected value maximization) is proposed. This method can detect and eliminate global and local outliers of the joint time-domain feature sets, which makes the classification more reasonable.

(3) The pattern recognition method of CGF is proposed. The method can determine the number of categories of FCM adaptively by cyclic GMM-EM algorithm, which makes the classification result more accurate.

(4) The effectiveness and superiority of joint time-domain features based on VKF as feature sets are verified by the gearbox vibration experiment under various states.

(5) The effectiveness and superiority of CGF pattern recognition algorithm based on joint time-domain features are verified by the gearbox vibration experiments in various states.

II. METHODOLOGY

A. PRINCIPLE OF GMM-EM AND ANOMALY DETECTION

Gaussian mixture model (GMM) [19] is an extension of the Gaussian model and a linear combination of several Gaussian distribution functions. GMM assumes that all sample data obey the mixed Gaussian distribution, that is, the probability density function of sample data sets is estimated [20]. The estimated model is the linear combination of the Gaussian model, and each Gaussian distribution is a cluster. The clustering result of the hybrid model is that data sets are divided into several clusters which obey the independent Gaussian distribution function [21] based on probability.

Suppose the random variable is X, and the mixing model is composed of Gaussian distributions whose number is M. GMM can be expressed as:

$$P(x|\theta) = \sum_{m=1}^{M} \alpha_m \phi(x|\theta_m) \tag{1}$$

where, the parameter α_m represents the weight of the m-th Gaussian distribution in GMM and satisfies $\sum_{m=1}^{M} \alpha_m = 1, (\alpha_m \geq 0)$;

$\phi(x|\theta_m)$ is the probability density function of m-th Gaussian distribution. At the same time, θ_m expressed as: $\theta_m = (\mu_m, \Sigma_m)$, μ_m represents the mean of the m-th Gaussian distribution, Σ_m represents the covariance matrix of the m-th Gaussian distribution.

$\phi(x|\theta_m)$ can be expressed as:

$$\phi(x|\theta_m) = \frac{1}{\sqrt{2\pi\,\Sigma_m}} \exp\left(-\frac{(x-\mu_m)^2}{2\Sigma_m}\right) \tag{2}$$

This paper describes the three parameters of the GMM for θ, namely the mixed coefficient α_m, the mean μ_m and the covariance matrix Σ_m. To divide the sample data set into several clusters, it is necessary to know parameters' values of the Gaussian distribution that each part of the data in this data set obeys, and make the clustering results fit the observed data as much as possible. Among them, the observed data in the mixed model refers to the data are known in the data set to obey the known Gaussian distribution [22]. At the same time, the sample data set is also called the complete data, containing the observed random data $X = \{x_1, x_2, \ldots, x_N\}$ and the unobserved random variable $Z = \{z_1, z_2, \ldots, z_N\}$.

Thus, GMM is mainly determined by the parameter θ. In order to obtain a high quality clustering result, the optimal sample parameters need to be solved. The most common method is to maximize the logarithmic likelihood function of the mixed model. Expected value maximization (EM) [23] algorithm iteratively solves the parameters of the maximum likelihood estimation from complete data containing implicit variables.

The iteration of EM algorithm is accomplished by two major steps, Expectation step (E-Step) and Maximization Step(M-step) [24]. Because the goal is to solve the distribution parameters of GMM, and the implicit and unobserved data is unknown, the EM algorithm of E-step first guess the implicit data of the model to get the expected value and the Gaussian distribution [25]. Then maximum likelihood estimation is performed for complete data and the parameters of GMM are solved, namely M-step. However, in the process of solving the parameters of GMM, the implied data of E-Step is obtained by guessing [26]. It is not accurate so that the model parameters solved according to the results are not accurate. Therefore, the above steps are repeated again [27]. In this way, E-step and M-step are iterated continuously until the parameters of GMM are basically unchanged. The algorithm converges to find the optimal expectation, covariance matrix of GMM and weight of each Gaussian distribution.

According to Bayes' theorem [28], [29], by selecting the initial value of parameters of GMM, the influence degree of the m-th Gaussian distribution on the observed data $(x_1, x_2, \ldots, x_N)$ in GMM, namely, the maximum posterior probability is estimated as follows:

$$\hat{\gamma}_{jm} = \alpha_m \phi(x_j|\mu_m, \sigma_m^2) \Big/ \sum_{m=1}^{M} \alpha_m \phi(x_j|\mu_m, \sigma_m^2) \tag{3}$$

According to the initial value of selected model parameters, the expected value of the logarithmic likelihood function of the mixed model can be expressed as:

$$E_Q\left[\log p(\theta|Y, Q)|\theta^{(i)}, Y\right] = \int \log\left[p(\theta|Y, Q)\right] p(Q|\theta^{(i)}, Y) dQ \tag{4}$$

where, Q represents the implicit data, $\theta^{(i)}$ is the posterior standard deviation of the i+1 iteration.

The conditional expectation probability of the joint distribution of the mixed model can be expressed as:

$$L(\theta, \theta_j) = \sum_{i=1}^{m} \sum_{z_j} P\left(z_j|x_i, \theta_j\right) \log P\left(x_i, z_j|\theta\right) \tag{5}$$

Under the conditional probability constraint of E-step, the maximum value of logarithmic likelihood function parameters can be expressed as:

$$\theta_{j+1} = \arg\max_{\theta} L(\theta, \theta_j) \tag{6}$$

After a new round of iteration, the three parameters can be expressed as:

$$\begin{aligned}
\hat{\mu}_m &= \sum_{j=1}^{N} \hat{\gamma}_{jm} x_j \Big/ \sum_{j=1}^{N} \hat{\gamma}_{jm}, \quad m = 1, 2, \ldots, M \\
\hat{\Sigma}_m &= \sum_{j=1}^{N} \hat{\gamma}_{jm} \left(x_j - \mu_m\right)^2 \Big/ \sum_{j=1}^{N} \hat{\gamma}_{jm}, \quad m = 1, 2, \ldots, M \\
\tilde{\alpha}_m &= \sum_{j=1}^{N} \hat{\gamma}_{jm} \Big/ N, \quad m = 1, 2, \ldots, M
\end{aligned} \tag{7}$$

The above E-step and M-step are iterated until values of $\theta^{(i)}$ and $\theta^{(i+1)}$ are infinitely close to each other.

After obtaining the final convergent GMM, the generation probability F_k of each sample point x_k can be expressed as:.

$$F_k = \sum_{i}^{m} \alpha_i p\left(x_k|\mu_i, \sigma_i\right) \tag{8}$$

F_k is lower, the more likely x_k is to be an outlier. Therefore, in GMM algorithm, the derivative of F_k is taken as the score to determine the anomaly.

B. PRINCIPLE OF FCM

FCM (Fuzzy C-means) algorithm is a classical algorithm based on distance. The distance value between the data and the cluster class center is calculated as the evaluation standard [17]. If they are more similar, the smaller the value is, and the greater probability of being divided into the same category is.

FCM algorithm is a typical clustering algorithm based on objective function, and uses Euclidean distance as similarity discrimination [30], [31]. It uses constraints to solve of objective function. Finally, the final solution is obtained through continuous iteration, and different types of data are divided into different cluster classes. Its definition is as follows [18]:

Suppose $X = \{x_1, x_2, \ldots, x_n\}$ is a set of feature data sets in space, $x_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ is a data sample in X, representing a point in the space sets, and x_{ij} is the j-th attribute value of x_i.

The clustering of a given data sets X is to generate c classes of $X_1, X_2, \ldots, X_c$. The membership degree μ_{ik} represents the membership relationship between sample x_i and subset X_k.

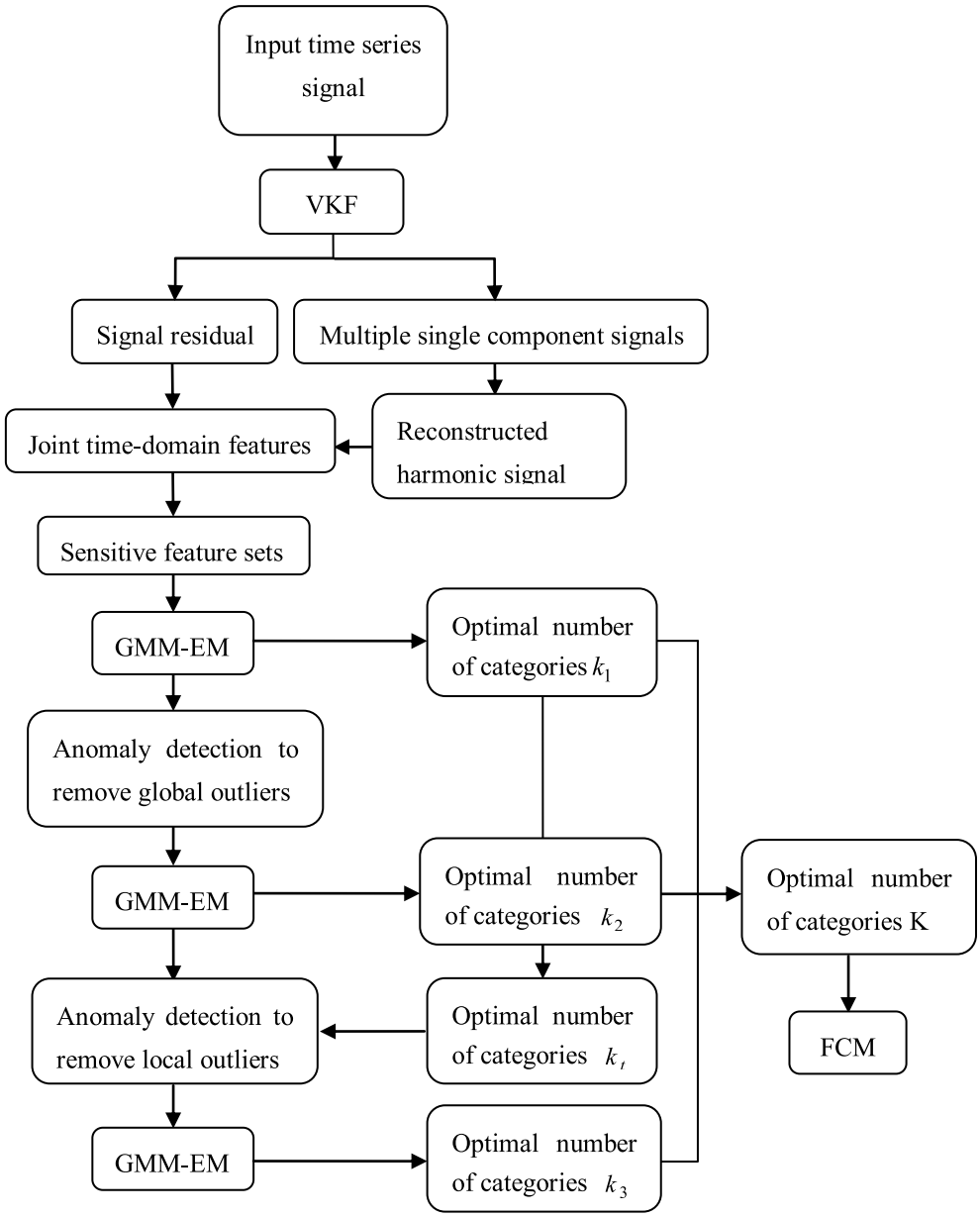


FIGURE 1. The algorithm flowchart of cyclic GMM-FCM pattern recognition based on joint time-domain features.

It satisfies $\mu_{ik} \in [0,1] \,\&\forall i, \sum_{k=1}^{c} \mu_{ik} = 1; \forall k, 0 < \sum_{k=1}^{c} \mu_{ik} < n, U = (\mu_{ik})$ is $n \times c$ dimension of membership function.

The objective function of FCM is expressed as:

$$J = (U, V) = \sum_{k=1}^{c} \sum_{i=1}^{n} (\mu_{ik})^m d_{ik}^2 \tag{9}$$

where, $d_{ik} = \|x_i - V_k\|$ refers to the Euclidean distance between x_i and the k-th clustering center V_k; $m \in (1, \infty)$ is the fuzzy weighted exponent; $V = (V_1, V_2, \ldots, V_c)$ is the cluster class center set of all subsets X_k.

Combined with Lagrange multiplication $\sum_{k=1}^{c} \mu_{ik} = 1, \mu_{ik} \in [0,1], i = 1, 2, \ldots, n$, available:

$$\sum_{k=1}^{c} \mu_{ik} = \left[\sum_{j=1}^{c} \left(d_{ik} / d_{ij} \right)^{2/(m-1)} \right]^{-1} \tag{10}$$

$$V_k = \sum_{i=1}^{n} (\mu_{ik})^m x_i \Big/ \sum_{i=1}^{n} (\mu_{ik})^m \tag{11}$$

FCM algorithm steps are as follows:

(1) Given c, m, ε and initial iteration value $f = 0$, U is initialized by using random number between [0, 1], satisfy the constraint conditions $\sum_{k=1}^{c} \mu_{ik} = 1, \forall i = 1, 2, \ldots, n$;

(2) According to Equation (11), cluster centers V_k are calculated and the number is c;

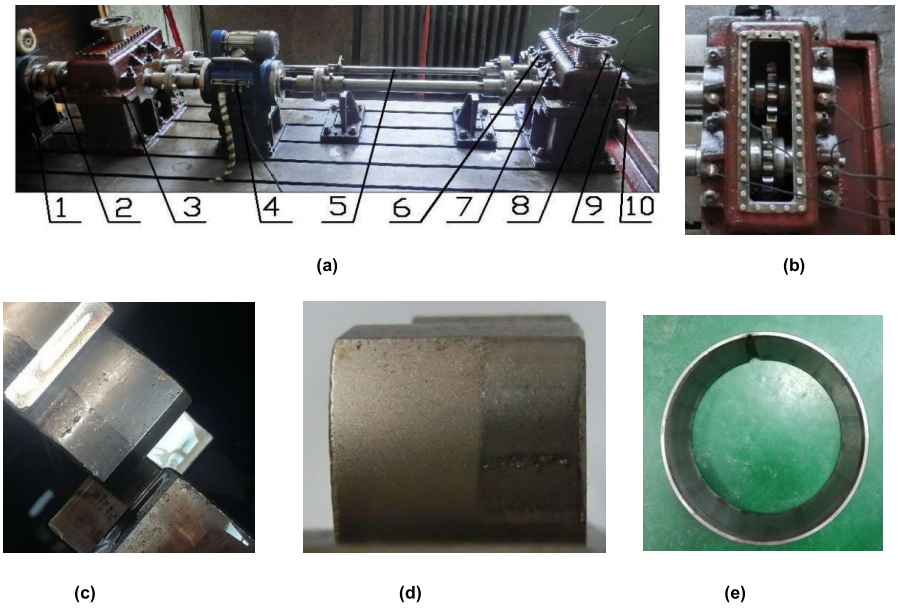


FIGURE 2. (a) is the test bench, where, 1-speed motor, 2-coupling, 3-test gearbox, 4-torque tachometer, 5-torque bar, 6/7/8/9-piezoelectric acceleration sensor, 10-main test gearbox; (b) is the schematic diagram of sensors distribution; (c) is slight gear pitting; (d) is gear pitting; (e) is electric discharge machined bearing outer ring.

TABLE 1. Independent features and combined features.

Independent features	variance; mean square value; root mean square value; skewness; kurtosis; waviness index; margin index; pulse index; peak index ; kurtosis index		
Combined features	Time domain & frequency domain	Time domain	Frequency domain
		mean; variance; variance coefficient; skewness; kurtosis	spectral flatness; spectral entropy; spectral flux
	Time-frequency domain	mean; variance; variance coefficient; skewness; kurtosis; time -frequency entropy flatness; Shannon entropy; time-frequency flux	

(3) The $U^{(f)}$ of this iteration is calculated according to Equation (10), and the $J^{(f)}$ of this iteration is calculated according to Equation (9);

(4) If the condition $\left|J^{(f)} - J^{(f-1)}\right| \leq \varepsilon$ is true, then output the result; otherwise, $f = f + 1$, turn step (2).

C. ALGORITHM FLOW CHART AND MAIN STEPS

The algorithm flow chart of cyclic GMM-FCM based on joint time-domain features is designed by the above theory, and the results are shown in Figure 1.

Where, VKF can separate complex multi-component signals [32] into a combination of multiple single-component signals and signal residues through the instantaneous frequency curve of each component in the obtained signal. The single component signal is related to the parts that produce vibration impact [33], while the signal residue contains noise and hidden information [34]. The combination provides the possibility to extract the features of different components [35].

The main steps of the improved algorithm of pattern recognition are as follows:

(1) An improved signal separation method of VKF is used to decompose the input signal into harmonic reconstruction signal and signal residual. Extract the time-domain features as a dimension to build 2D joint time domain. The joint time-domain feature sets of signals in different states are regarded as feature sets. Sensitive feature sets are selected from ten types of joint time-domain feature sets;

(2) The pattern recognition algorithm conducts preliminary training on the feature set with GMM-EM algorithm. It obtains the initial optimal number of categories k_1, and removes the global outliers in the feature set through GMM anomaly detection;

(3) GMM-EM training is carried out on the feature set again to obtain the optimal number of categories k_2. The optimal category number k_t was obtained by rounding the mean values of k_1 and k_2. The feature set is preliminarily labeled by k_t.

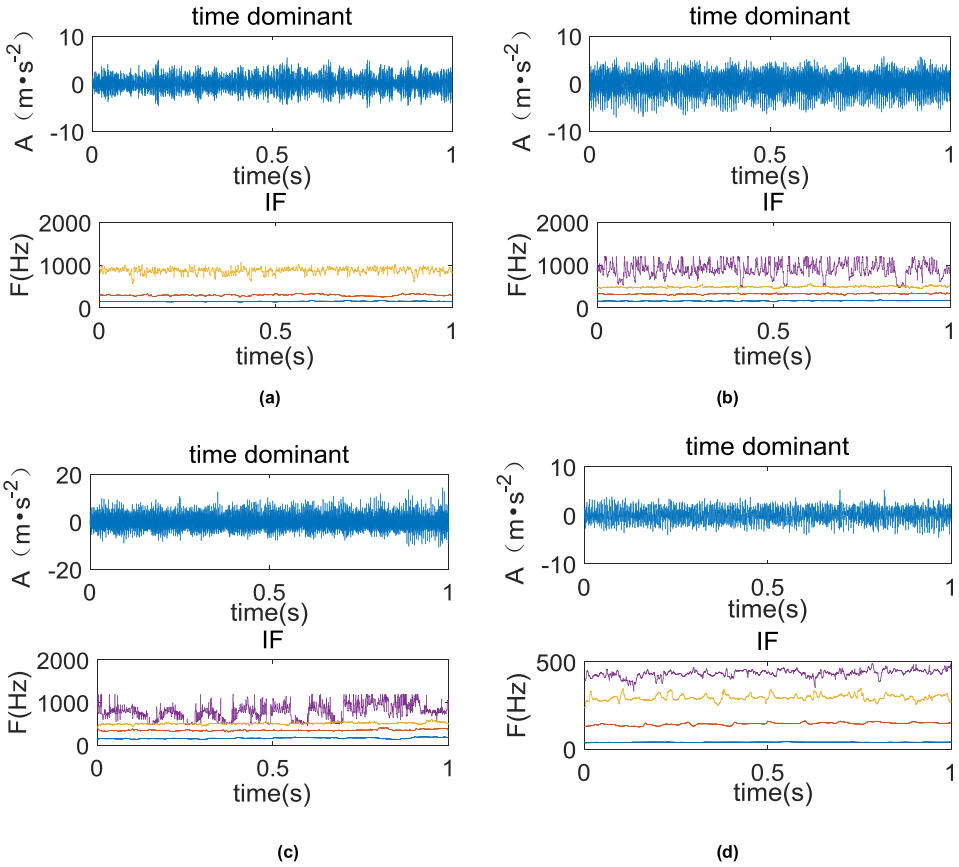


FIGURE 3. (a) is the normal signal, (b) is the slight gear-pitting signal, (c) is the gear-pitting signal and (d) is the signal of composite faults.

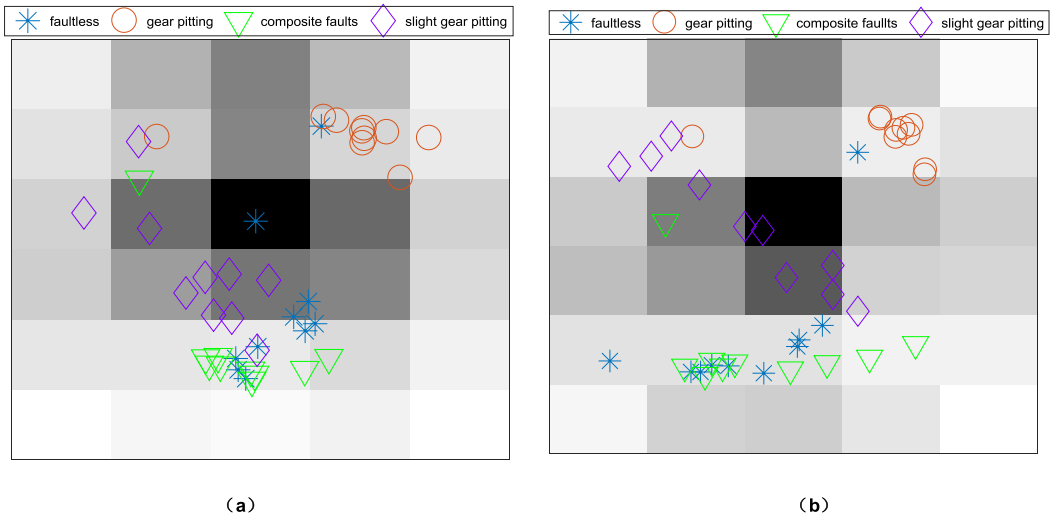


FIGURE 4. Training results of GTM, where, (a) is 20 × 40 combined feature sets after GTM and (b) is 28 × 40 combined feature sets after GTM.

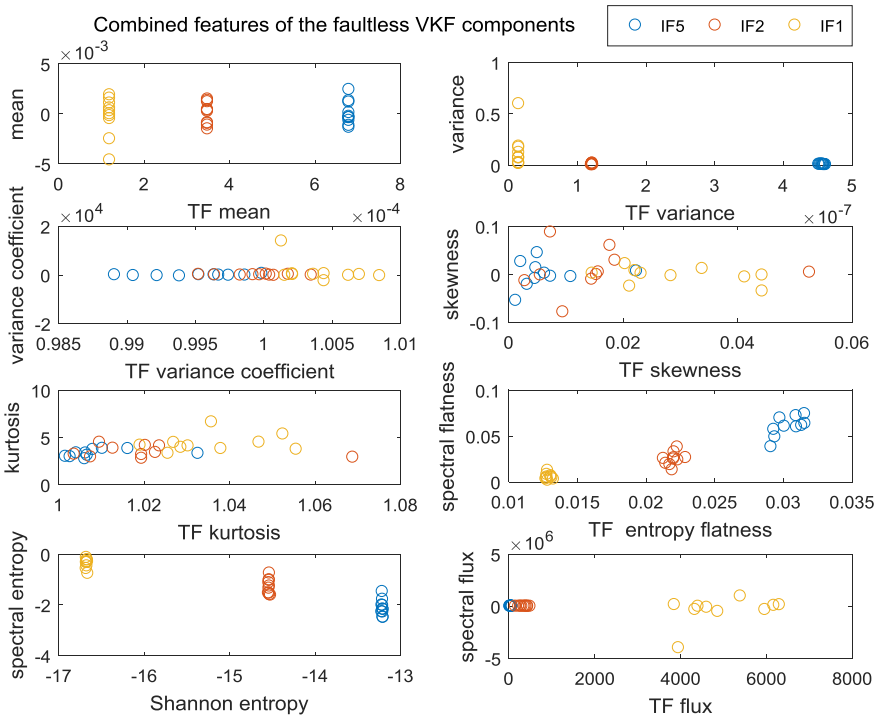
(4) GMM anomaly detection is carried out for each type of label feature data to remove local anomaly points. Then GMM-EM training is carried out on the feature set again to obtain the optimal number of categories k_3.

(5) The optimal number of categories K was obtained by roundness of the mean values of k_1, k_2 and k_3, and K was taken as the category number of FCM to carry out clustering training for feature sets with global and local anomalies removed.

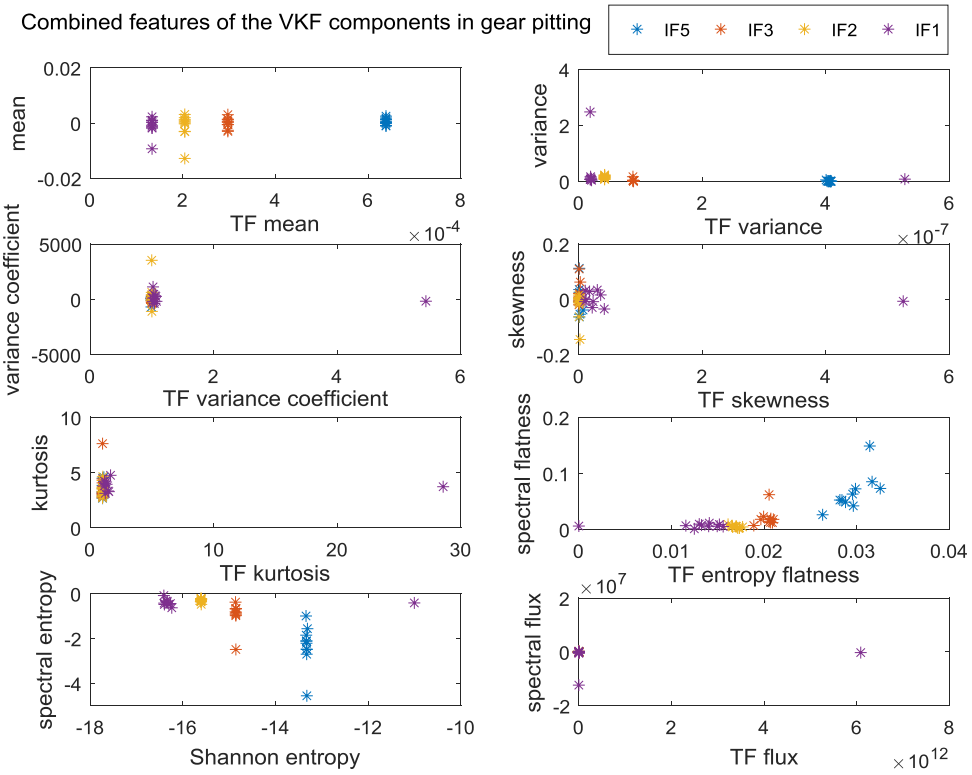
III. SELECTION OF JOINT TIME-DOMAIN FEATURES OF COMPONENT SIGNALS

A. THE DESIGN SCHEME OF EXPERIMENT

As shown in Table 1, in this paper, ten time-domain features, such as variance, mean square value, root mean square

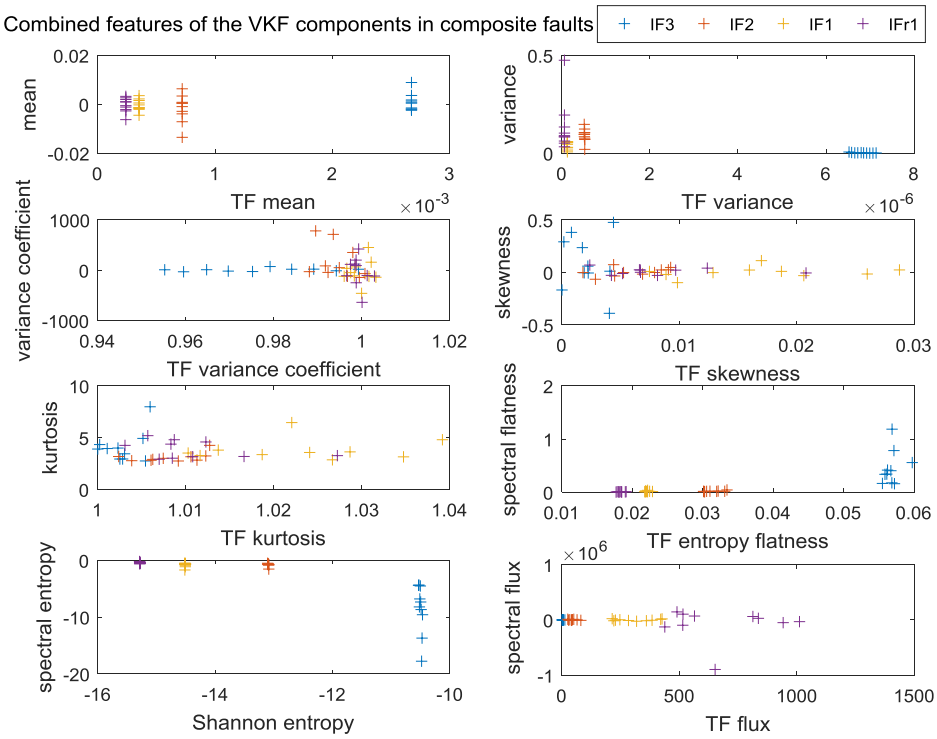


(a)



(b)

FIGURE 5. Combined features of VKF components in different states, where, (a) is in the faultless state, (b) is in the state of gear pitting and (c) is in the state of composite faults.



(c)

FIGURE 5. *(Continued.)* **Combined features of VKF components in different states, where, (a) is in the faultless state, (b) is in the state of gear pitting and (c) is in the state of composite faults.**

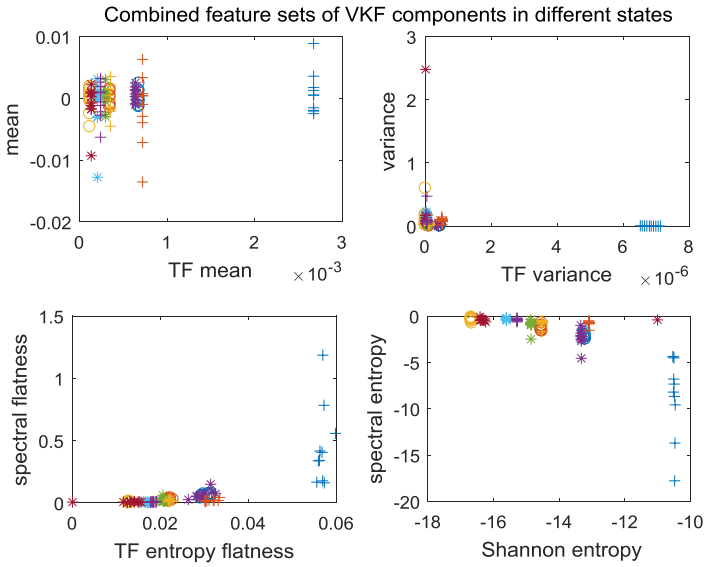
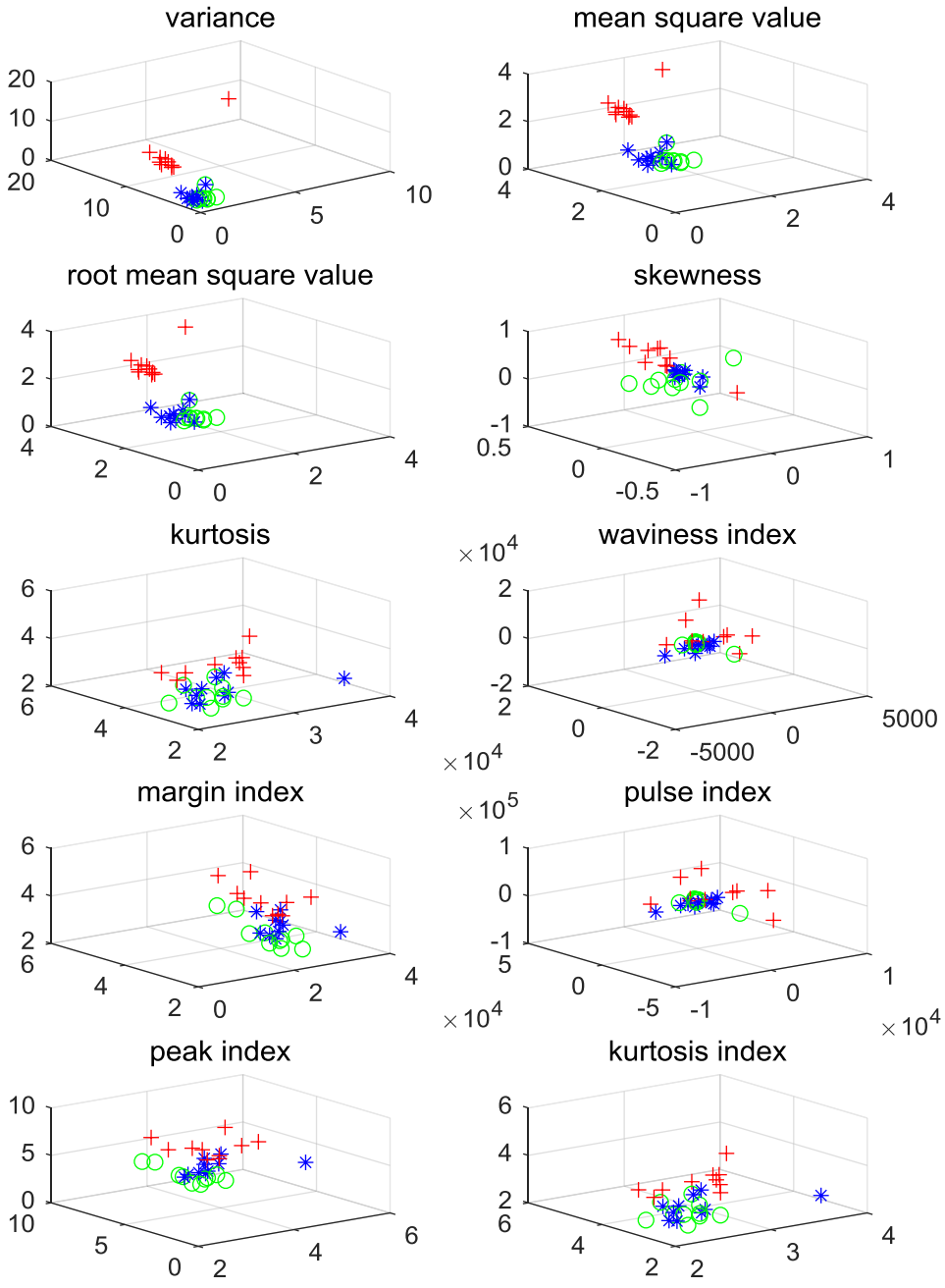


FIGURE 6. **Combined feature sets of VKF components in different states.**

value, skewness, kurtosis, waviness index, margin index, pulse index, peak index and kurtosis index, are selected as independent features. Select time-domain features of mean, variance, variance coefficient, skewness and kurtosis, frequency-domain features of spectral flatness, spectral entropy and spectral flux, and time -frequency domain features of mean, variance, variance coefficient, skewness, kurtosis, time -frequency entropy flatness, Shannon entropy and time-frequency flux, such as eight group features, to constitute combined features that the time-frequency domain features correspond to the time-domain and frequency-domain features.

In order to verify the practicability of the improved method proposed in this paper, a closed power flow gearbox test

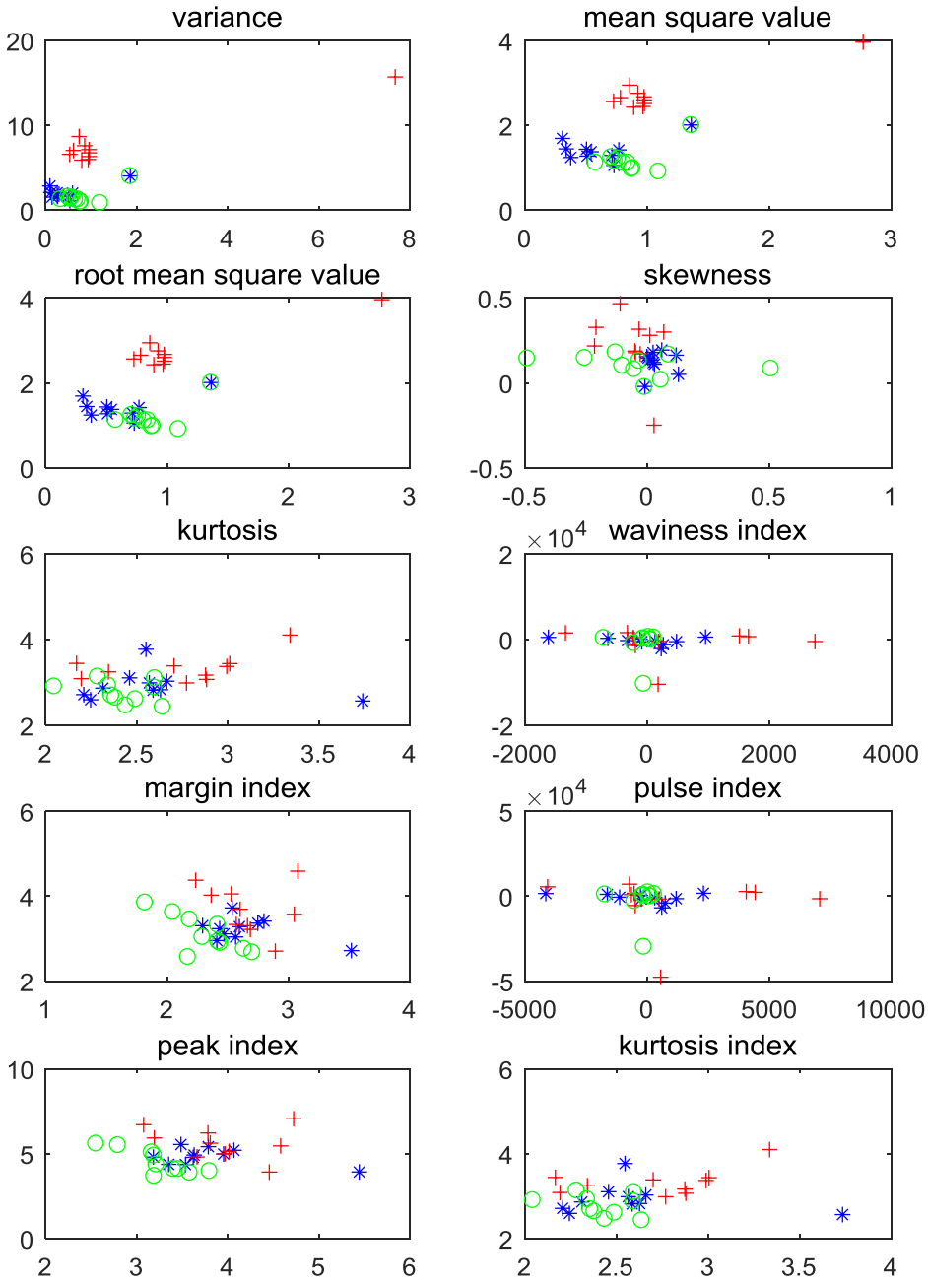


(a) 3D joint time-domain feature sets, where, features of reconstructed signals are the X-axis, features of signal residuals are the Y-axis, and features of original signals are the Z-axis

FIGURE 7. Joint time-domain feature sets.

bench was used to perform related experiments. The vibration signals of the gearbox were collected under normal, gear pitting transition, gear pitting and combined faults (gear pitting and bearing outer ring pitting). The bench is loaded by the internal force generated by the torsion bar. The test gear has a transmission ratio of 1: 1 and a number of 18 teeth. The test bench is shown in Figure 2 (a). The model of piezoelectric sensor is CA-YD-186 (sensitivity is 10.41mV/m·s^2), the sampling frequency is 12000Hz, and layout positions are shown in Figure 2 (b). Figure 2 (c) (d) (e) show the parts in three fault states.

The signals in different states are selected, as shown in Figure 3(a) (b) (c) (d), which are the normal signal, the slight gear-pitting signal, the gear-pitting signal and the signal of



(b) 2D joint time-domain feature sets, where, features of reconstructed signals are the X-axis and features of signal residuals are the Y-axis

FIGURE 7. ***(Continued.)*** **Joint time-domain feature sets.**

combined faults. They have a load of 800N · M and a uniform acceleration of 550r/min to 600r/min.

In view of the multi-dimensional features of signals, the general method is to construct high dimensional sets, and to find hidden variables in the high-dimensional feature sets through feature mapping (such as principal component analysis, self-organizing mapping, etc.), so as to classify feature sets and realize the pattern recognition of fault types.

As shown in Figure 4, (a) combines 10 independent features of reconstructed signals and signal residuals to construct a 20 × 40 high-dimensional feature sets, and (b) combines 8 groups of combined features and 10 independent features of reconstructed signals and signal residuals to construct a 28 × 40 high-dimensional feature sets. After the labels of feature sets in (a) and (b) are marked, the sets are classified by generative topographic mapping (GTM). The results show that the dimension of feature sets is higher, the

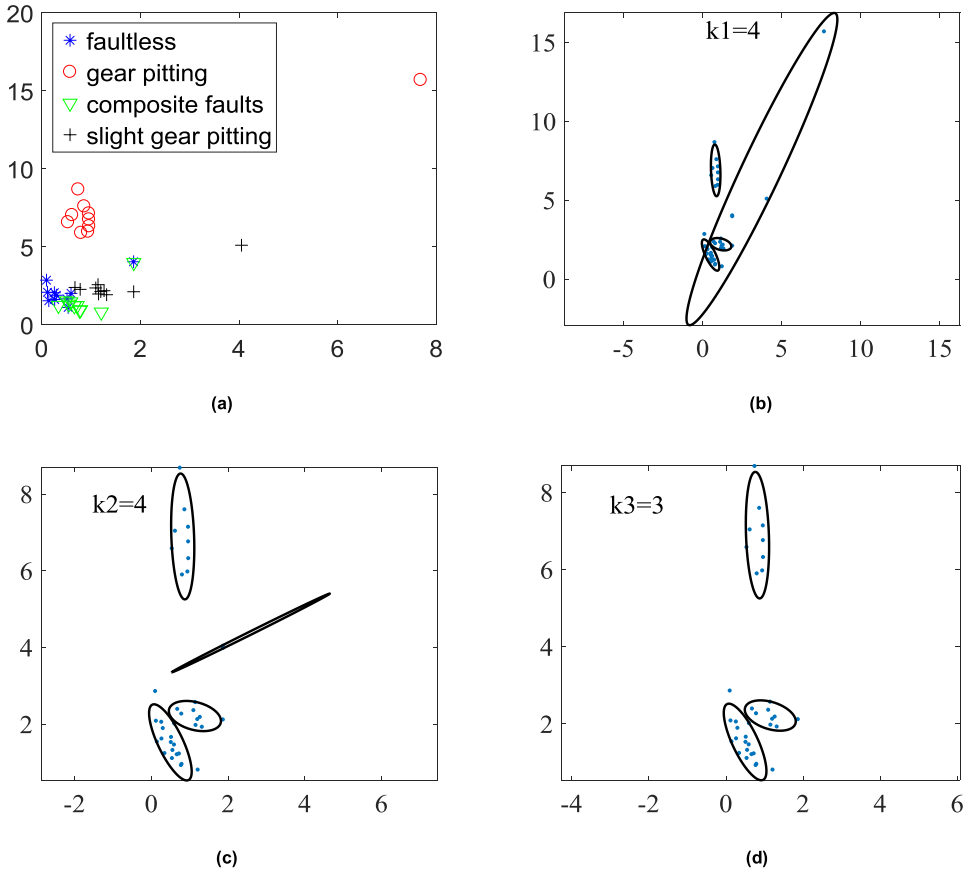


FIGURE 8. The training results of the cyclic GMM-EM algorithm for sensitive feature sets, where, (a) is sensitive feature sets, (b) is the result of preliminary GMM-EM training, (c) is the result of GMM-EM training after global outliers removed and (d) is the result of GMM-EM training after local outliers removed.

result of pattern recognition is better. Although it can carry on the classification to different fault states, it has to mark labels in advance. In the mapping results of the feature sets, there are outliers far away from the cluster of each fault state. And due to the problem such as ''dimension disaster'', it is difficult to eliminate abnormal feature sets.

Therefore, it is necessary to preserve the integrity of signal information in the low-dimensional data set as much as possible and to show the relationships hidden in the high-dimensional data set.

B. ANALYSIS OF COMBINED FEATURES AND JOINT TIME-DOMAIN FEATURE SETS

Previous studies have shown that the feature set composed of single combined feature or independent feature cannot be used for pattern recognition. Therefore, this paper attempts to start with characteristics of VKF components and conduct research in two ways. On the one hand, 8 pairs of combined features are taken as horizontal and vertical coordinates, and feature sets are constructed with the combined eigenvalues of each component of signals in different states. Then, 8 groups of feature sets are screened to select sensitive feature sets for pattern recognition. On the other hand, reconstructed harmonic signals of VKF, signal residues and original signals are selected as dimensions, and their time-domain independent features are taken as values to construct a two-dimensional or three-dimensional joint time-domain feature set.

Combined features are extracted for each component signal and signal residual of signals after VKF. Ten sets of gear box vibration signals are selected for each state, and the results are shown in Figure 5.

According to Figure 5(a), (b) and (c), the eight combination features of VKF components in different states, the four combination features of group 1, group 2, group 6 and group 7 can better distinguish each component, namely, the combined features of the four groups of mean-TF mean, variance-TF variance, spectral flatness-TF entropy flatness and spectral entropy-Shannon entropy.

Each VKF component in different states is combined into a feature set according to the above four combined features, as shown in Figure 6. The results show that the feature set composed of each component has no clear boundary in vision, and there are abnormal points in some combination features, so it is difficult to carry out clustering analysis.

Therefore, the idea that constructing a feature set from combined features of each component is abandoned. Independent features of reconstructed signals, signal residues and original signals are obtained respectively to construct the feature set, which is named as joint time-domain feature set.

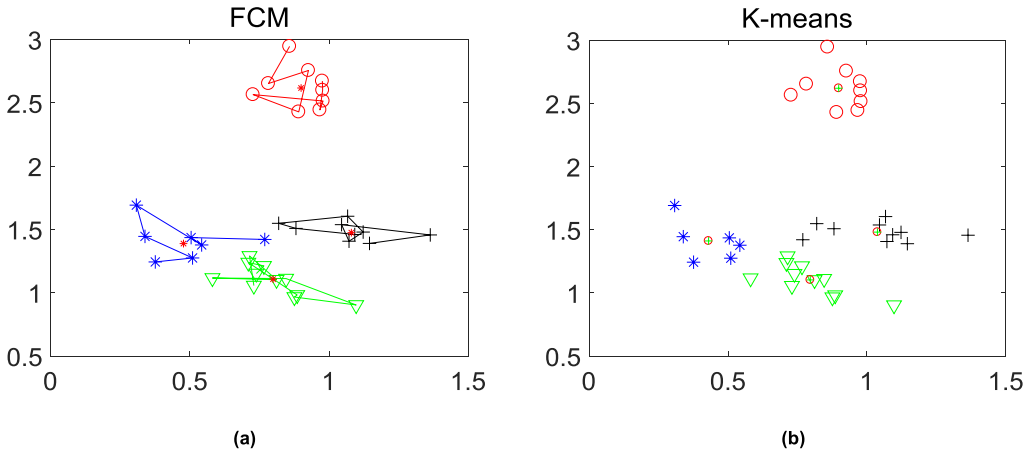


FIGURE 9. (a) is FCM after cyclic GMM-EM and (b) is k-means after cyclic GMM-EM.

Independent features are ten independent features listed in Table 1.

Reconstructed harmonic signals of VKF, signal residues and original signals are selected as dimensions, and their time-domain independent features are taken as values to construct a two-dimensional or three-dimensional joint time-domain feature set.

As shown in Figure 7, abnormal points both exist in the 2D and 3D, and the joint time-domain feature set in (a) is more clear than (b) to distinguish the boundaries of different states. But due to anomaly detection of 2D feature set is the most simple, and the sum of reconstructed signal and residual is the original signal, that is, 2D feature set contains all the features of vibration state information. 3D feature set increase dimension on the basis of the original signal characteristics. It makes part of the information redundancy. Therefore, this paper chooses the two-dimensional joint time-domain features to construct the feature set.

IV. EXPERIMENTAL VERIFICATION

In Figure7 (b), 1 to 3 groups of sensitive joint features, boundary of different state is obvious, and the distribution of three groups of joint features are similar. Therefore, Joint time-domain feature set of variance is chosen as the sensitive feature set. And train the sensitive feature set with the improved algorithm of pattern recognition in this paper. The results are shown in Figure 8.

$k_1 = 4$ is obtained by preliminary GMM-EM training on the feature set in Figure 8 (a), and the preliminary classification result is shown in (b). It shows that the classification result is greatly affected by abnormal points. After removing the global outliers, the GMM-EM training result is shown in (c) and $k_2 = 4$. Compared with (b), the classification result of (c) is better. But it is still greatly affected by local outliers. The optimal number of categories $k_t = 4$ was obtained by rounding the mean values of $k_1$and k_2. After average classification of feature sets, local outliers are removed. And the result of GMM-EM training is shown in (d) and $k_3 = 3$. It can be found that not all the points in (d) are in the classification results. It indicates that the GMM-EM algorithm, as a hard clustering algorithm, cannot consider all feature points.

By rounding the mean values of k_1, k_2 and k_3, the optimal number of categories K=4 is obtained. The feature sets with global and local outliers removed were trained by fuzzy C-means clustering (FCM) and K-means clustering, respectively. The results are shown in Figure 9 (a) and (b). Compared with Figure 8(a), the training accuracy of FCM is better than that of K-means clustering. It achieves the purpose of relatively accurate pattern recognition and lays a solid foundation for subsequent fault diagnosis.

V. CONCLUSION AND RECOMMENDATIONS

This paper proposes a pattern recognition method of cyclic GMM-FCM based on joint time- domain features. In this paper, according to the characteristics of VKF components, a number of combined features and independent features are selected to test. The results show that the sensitive feature set constructed by joint time-domain features can retain all the information of signal features while avoiding the "dimensional disaster" problem faced by anomaly detection. Then, the pattern recognition method based on cyclic GMM-FCM can effectively eliminate global and local outliers in the joint time-domain feature set, and determine the number of FCM categories adaptively. It makes the result of classification more accurate and lays a foundation for the application of the improved method in the industrial environment.

In future work, it is recommended to extend this improved method to other composite faults and vibration environments, such as composite faults of planetary gearboxes.

REFERENCES

[1] W. Chen, Z. Huang, F. Wu, M. Zhu, and H. Guan, "VAUD: A visual analysis approach for exploring spatio-temporal urban data," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 9, pp. 2636–2648, Sep. 2018.

[2] H. Wang, W. Song, E. Zio, A. Kudreyko, and Y. Zhang, "Remaining useful life prediction for lithium-ion batteries using fractional Brownian motion and fruit-fly optimization algorithm," *Measurement*, vol. 161, Sep. 2020, Art. no. 107904.
[3] J. Xia, Y. Hou, Y. V. Chen, Z. C. Qian, D. S. Ebert, and W. Chen, "Visualizing rank time series of wikipedia top-viewed pages," *IEEE Comput. Graph. Appl.*, vol. 37, no. 2, pp. 42–53, Apr. 2017.
[4] Z. He, H. Shao, P. Wang, J. J. Lin, and J. Cheng, "Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowl.-Based Syst.*, vol. 191, Mar. 2020, Art. no. 105313.
[5] N. Cao, Y. R. Lin, D. Gotz, and F. Du, "Z-Glyph: Visualizing outliers in multivariate data," *Inform. Vis.*, vol. 17, no. 1, pp. 22–40, Jan. 2018.
[6] S. McKenna, D. Staheli, and C. Fulcher, "Bubblenet: A cyber security dashboard for visualizing patterns," *Comput. Graph. Forum*, vol. 35, no. 3, pp. 281–290, Jun. 2016.
[7] T. Zhang, X. Wang, and Z. Li, "A survey of network anomaly visualization," *Sci. China Inf. Sci.*, vol. 60, no. 12, pp. 126–142, 2017.
[8] H. Fanaee-T and J. Gama, "Tensor-based anomaly detection: An interdisciplinary survey," *Knowl.-Based Syst.*, vol. 98, pp. 130–147, 2016.
[9] J. Liu and Y. Shao, "Overview of dynamic modeling and analysis of rolling element bearings with localized and distributed faults," *Nonlinear Dyn.*, vol. 93, no. 4, pp. 1765–1798, 2018.
[10] M. Pradhan and P. Guota, "Fault detection using vibration signal analysis of rolling element bearing in time domain using an innovative time scalar indicator," *Int. J. Manuf. Res.*, vol. 12, no. 3, pp. 305–317, 2017.
[11] K. Taegong and P. Cheong Hee, "Anomaly pattern detection for streaming data," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113252.
[12] Z. He, H. Shao, X. Zhang, J. Cheng, and Y. Yang, "Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples," *IEEE Access*, vol. 7, pp. 115368–115377, 2019.
[13] L. He, W. Song, M. Li, A. Kudreyko, and E. Zio, "Fractional Lévy stable motion: Finite difference iterative forecasting model," *Chaos, Solitons Fractals*, vol. 133, no. 4, 2020, Art. no. 109632.
[14] H. Shen, J. Chen, R. Wang, and J. Zhang, "Counterfeit anomaly using generative adversarial network for anomaly detection," *IEEE Access*, vol. 8, pp. 133051–133062, 2020.
[15] G. Wang, J. Yang, and R. Li, "Imbalanced SVM-based anomaly detection algorithm for imbalanced training datasets," *ETRI J.*, vol. 39, no. 5, pp. 621–631, Oct. 2017.
[16] H. Wang, J. Gu, and S. Wang, "An effective intrusion detection framework based on SVM with feature augmentation," *Knowl.-Based Syst.*, vol. 136, pp. 130–139, Nov. 2017.
[17] S. Askaria, N. Montazerina, and M. Zarandib, "Generalized possibilistic fuzzy C-means with novel cluster validity indices for clustering noisy data," *Appl. Soft Comput.*, vol. 53, pp. 262–283, 2017.
[18] H. Chen, X. Shen, and J. Long, "A novel automatic fuzzy clustering algorithm based on soft partition and membership information," *Neurocomputing*, vol. 236, pp. 104–112, 2017.
[19] J. Liu, Z. Ren, R. Lu, and X. Luo, "GMM discriminant analysis with noisy label for each class," *Neural Comput. Appl.*, Jun. 2020.
[20] P. Sandeep and T. Jacob, "Single image super resolution using a joint GMM method," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4233–4244, Jul. 2016.
[21] Q. Jiang, B. Huang, and X. Yan, "GMM and optimal principal components-based Bayesian method for multimode fault diagnosis," *Comput. Chem. Eng.*, vol. 84, pp. 338–349, Jan. 2016.
[22] Y. Wang, D. Yang, X. Zhang, and Z. Chen, "Probability based remaining capacity estimation using data-driven and neural network model," *J. Power Sources*, vol. 315, pp. 199–208, May 2016.
[23] B. Zhao, K. Setsompop, H. Ye, S. F. Cauley, and L. L. Wald, "Maximum likelihood reconstruction for magnetic resonance fingerprinting," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1812–1823, Aug. 2016.
[24] V. Christlein, D. Bernecker, F. Hönig, A. Maier, and E. Angelopoulou, "Writer identification using GMM supervectors and exemplar-SVMs," *Pattern Recognit.*, vol. 63, pp. 258–267, Mar. 2016.
[25] A. Meng, J. Yang, M. Li, and S. Jiang, "Research on hysteresis compensation control of GMM," *Nonlinear Dyn.*, vol. 83, nos. 1–2, pp. 161–167, 2016.
[26] B. Celler, P. Le, A. Argha, and E. Ambikairajah, "GMM-HMM based blood pressure estimation using time-domain features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3631–3641, Aug. 2019.
[27] P. Sengottuvelan and N. Prasath, "BAFSA: Breeding artificial fish swarm algorithm for optimal cluster head selection in wireless sensor networks," *Wireless Pers. Commun.*, vol. 94, no. 4, pp. 1979–1991, 2017.
[28] Z.-Y. Chen and R. J. Kuo, "Combining SOM and evolutionary computation algorithms for RBF neural network training," *J. Intell. Manuf.*, vol. 30, no. 3, pp. 1137–1154, Mar. 2019.
[29] J. de Jesús Rubio, "Least square neural network model of the crude oil blending process," *Neural Netw.*, vol. 78, pp. 88–96, Jun. 2016.
[30] X. Li, B. Li, F. Liu, H. Yin, and F. Zhou, "Segmentation of pulmonary nodules using a GMM fuzzy C-means algorithm," *IEEE Access*, vol. 8, pp. 37541–37556, 2020.
[31] J. Mao, C. Jin, and Z. Zhang et a1, "Anomaly detection for trajectory big data: Advancements and framework," *J. Softw.*, vol. 28, no. 1, pp. 17–34, 2017.
[32] H. Vold, M. Mains, and J. Blough, "Theoretical foundations for high performance order tracking with the Vold-Kalman tracking filter," *SAE Trans.*, vol. 106, pp. 3046–3050,1997.
[33] K. Feng, K. Wang, and M. Zhang et a1, "A diagnostic signal selection scheme for planetary gearbox vibration monitoring under non-stationary operational conditions," *Meas. Sci. Technol.*, vol. 28, no. 3, p. 1, 2017.
[34] P. Kundu, A. Darpe, and M. Kulkarni, "A correlation coefficient based vibration indicator for detecting natural pitting progression in spur gears," *Mech. Syst. Signal Process.*, vol. 129, pp. 741–763, 2019.
[35] K. Feng, W. A. Smith, P. Borghesani, R. B. Randall, and Z. Peng, "Use of cyclostationary properties of vibration signals to identify gear wear mechanisms and track wear evolution," *Mech. Syst. Signal Process.*, vol. 150, Mar. 2020, Art. no. 107258.

YANFENG LI was born in Taiyuan, Shanxi, China, in 1990. He received the bachelor's degree in engineering from the University of Jinan, in 2014, and the master's degree in engineering from the Taiyuan University of Technology, in 2017, where he is currently pursuing the Ph.D. degree with the College of Mechanical and Vehicle Engineering.

From 2014 to 2019, he was committed to the fault diagnosis research of rotating machinery. He participated in a number of National Natural Science Foundation of China. He has been responsible for the design and analysis of experiments and has written reports on them. He has published a number of research articles.

Dr. Li has been awarded the doctoral scholarship, in 2018 and 2019.

ZHIJIAN WANG was born in Zhengzhou, Henan, China, in 1985. He received the Ph.D. degree in engineering from the Taiyuan University of Technology, in 2015.

In 2018, he entered Xi'an Jiaotong University, where he holds a postdoctoral position. He is currently an Associate Professor and a Master's Supervisor with the School of Mechanical Engineering, North University of China. Since 2015, he has published more than 40 academic articles, among which the first author or corresponding author has published more than 20 SCI papers related to the subject, two papers cited by ESI, seven papers included by EI, more than 20 national invention patents applied by the first inventor, four patents authorized, and one monograph.

Prof. Wang has presided the National Natural Science Foundation of China and the Natural Science Foundation of Shanxi Province. He is a review expert of international famous SCI journals, such as *ISA Transactions*, *Measurement*, *Journal of Vibration and Control*, and IEEE Access.

TIANSHENG ZHAO was born in China, in 1978. He received the bachelor's degree in industrial automation from East China Jiaotong University, in 2001. He is currently a Senior Engineer with the Zhengzhou Mechanical and Electrical Engineering Research Institute. His research interests include mechanical and electrical control and fault diagnosis technology. He has won one first prize and one second prize of China Shipbuilding Industry Corporation Science and Technology Award.

SONG WANQING received the B.Sc. degree from the Inner Mongolia University of Science and Technology, in 1983, the M.Sc. degree from the University of Science and Technology Beijing, in 1990, and the Ph.D. degree from Donghua University, in 2010. He is currently a Professor with the Shanghai University of Engineering Science. His main research interests include condition monitor and fault diagnosis, bag date analysis, health, and reliability.