

Research Article

A Novel Clustering Method Combining ART with Yu's Norm for Fault Diagnosis of Bearings

Zengbing Xu,¹ Youyong Li,¹ Zhigang Wang,¹ and Jianping Xuan²

¹School of Mechanic and Automation, Wuhan University of Science and Technology, Wuhan, Hubei 430073, China

²School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430081, China

Correspondence should be addressed to Zengbing Xu; xuzengbing@163.com

Received 8 December 2015; Revised 4 April 2016; Accepted 18 April 2016

Academic Editor: Mariano Artés

Copyright © 2016 Zengbing Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Clustering methods have been widely applied to the fault diagnosis of mechanical system, but the characteristic that the number of cluster needs to be determined in advance limits the application range of the method. In this paper, a novel clustering method combining the adaptive resonance theory (ART) with the similarity measure based on the Yu's norm is presented and applied to the fault diagnosis of rolling element bearings, which can be adaptive to generate the number of cluster by the vigilance parameter test. Time-domain features, frequency-domain features, and time series model parameters are extracted to demonstrate the fault-related information about the bearings, and then considering the irrelevance or redundancy of some features many salient features are selected by an improved distance discriminant technique and input into the proposed clustering method to diagnose the faults of bearings. The experiment results confirmed that the proposed clustering method can diagnose the fault categories accurately and has better diagnosis performance compared with fuzzy ART and Self-Organizing Feature Map (SOFM).

1. Introduction

In order to decrease the downtime on production machinery and to increase reliability against possible failures, some important machinery is equipped with condition monitoring systems, but how to be intelligent to classify the data samples collected by the condition monitoring system is challenging. Artificial neural networks (ANNs) used as an intelligent classification tool have been widely applied in the fault diagnosis field of machine conditions which are treated as classification problems based on learning pattern from empirical data modeling in complex mechanical processes and systems [1]. But in the diagnosis process some ANNs are unable to detect unexpected fault changes. In other words, if the trained ANN has not learned a new fault category at the stage of training, the trained networks cannot identify it when the fault category occurs. In this case, these networks are required to retrain and learn the knowledge using the complete data sets; the trained network model must be modified to learn the new category. Thus, the previously learned knowledge is

forgotten and the memories of prior training are destroyed. This can result in a time-consuming and costly process. And in reality, it is impossible to obtain all the data set representing the features of all fault categories. These characteristics limit the application of these neural networks in fault diagnosis field. In order to solve this problem, the adaptive resonance theory (ART) network has been developed and applied to the field of pattern recognition and fault diagnosis. It is designed according to the adaptive resonance theory to overcome the stability-plasticity dilemma; that is, its learning system is able to protect useful historical data from the corruption (stability) while simultaneously learning new data (plasticity) [2, 3]. Owing to the advantages of ART network, some ART models such as ART2 and fuzzy ART have been developed and applied in some specified fields [4–6]. Furthermore, on the basis of advantage of ART some classification methods about the combination of ART and neural network, such as the CNN-ART algorithm [7] and ART-KOHONEN neural network [8], have been suggested to be applied in the field of fault diagnosis, which can be adaptive to expand the

knowledge continuously without the loss of the previous knowledge during learning new knowledge.

Currently clustering methods, owing to their superiority in independency of supervisors, have been widely studied and applied to the field of fault diagnosis. According to the principle that similar objects are within the same cluster and dissimilar objects are in different clusters, most of clustering methods mainly employ the similarity and distance measure to partition a dataset into several clusters, such as k nearest neighbor (KNN) and fuzzy c -means (FCM) [9–11]. Recently a novel clustering method using similarity measure derived from Yu's norm is also developed to classify medical datasets [12], which can deal with the uncertainty problem through the fuzzy formalism. But prior to classification, using these methods, the number of cluster nodes must be determined in advance. However, in the real world, it is rather hard to predict the number of classes. This is especially true in industry application where the mechanical equipment operates in a dynamic environment and under the influence of numerous uncertainty factors. Thus, these characteristics limit the application range of these clustering methods.

To the best of our knowledge, the clustering method using similarity measure based on the Yu's norm is seldom applied in the fault diagnosis of mechanical system. And in the fault diagnosis application when the samples of different fault classes overlap in some regions in the feature space, the traditional hard (crisp) clustering method mainly uses distance to compare data sample to fault classes and classify the data sample into one and only one cluster [13], these can result in the misclassification. But in fuzzy clustering the data sample is subject to one cluster with a certain grade of membership, and the clustering method based on similarity measure makes classification decision by comparing how similar the data sample is to the class vectors. Therefore, it is preferable to make use of the similarity measure based on the Yu's norm to develop a new diagnostic method which is capable of learning from the process data steam by identifying the different fault categories automatically; namely, the novel diagnostic method can be adaptive to generate the number of cluster nodes according to the number of faults in real time. Based on the elaboration mentioned above, a novel clustering method combined ART with similarity measure (ART-similarity) based on Yu's norm is proposed to diagnose the faults of rolling element bearing.

The rest of the paper is organized as follows. The review of adaptive resonance theory and clustering method using similarity measure based on Yu's norm is introduced in Section 2. Section 3 describes the proposed ART-similarity clustering method based on Yu's norm. Section 4 explains the diagnosis system using the proposed clustering method; namely, many feature parameters are extracted through the signal processing methods to depict the related-fault information, and some salient features are selected by the modified distance discriminant technique to be input into the ART-similarity clustering method to diagnose the faults of bearings. Finally, the conclusions about the proposed clustering method in the field of fault diagnosis are given in Section 5.

2. Review of ART and Clustering Method Using Similarity Measure Based on Yu's Norm

Adaptive resonance theory (ART) which is treated as a theory of human cognitive information processing was designed by Carpenter and Grossberg in 1976. ART network is an online learning system, and it mainly utilizes the self-organization to develop the stable and plastic clustering of input samples; namely, the ART network uses the vigilance test to resolve the stability-plasticity dilemma. In the learning process when a new sample is input into the ART network, it can attempt to categorize the sample by comparing it with the stored weight vectors of existing cluster node which represented a category. If the sample is similar to the existing categories and the match degree is greater than or equal to the vigilance value, the sample is classified into the specified category and the weight vectors of the corresponding cluster node are modified. Otherwise, a new category is created without affecting the existing memory. Its detailed dynamic character and algorithm can be seen in [4].

Because the fuzzy relation can be called a similarity relation [14], some clustering methods using similarity measure based on fuzzy set theory are developed. In 2007, a similarity measure based on Yu's norm is constructed and the corresponding clustering method is presented by Luukka [12]. Yu's norms, namely, Yu's T -norm and Yu's S -norm, in fuzzy logic, are written as follows, respectively:

$$T(x, y) = \max[0, (1 + \lambda)(x + y + 1) - \lambda xy], \quad (1)$$

$$S_n(x, y) = \min[1, x + y + \lambda xy], \quad (2)$$

in which $\lambda > 1$. According to the equivalence relations from T -norms and S -norms and negation [7], the similarity measure can be derived by the following:

$$S(x, y) = T(S_n(\bar{x}, y), S_n(x, \bar{y})). \quad (3)$$

Thus, the data samples can be clustered by the similarity measure.

The clustering algorithm is described as follows. Assume that a set X of objects need to be classified into N different classes C_1, \dots, C_N by D different feature parameters which depict the information about the object, and these feature parameters are normalized so that these objects are vectors that belong to $[0, 1]^D$. Then, the weight vector $V_i = (v_i(1), \dots, v_i(D))$ ($i = 1, \dots, N$) which presents the class i can be obtained by calculating the arithmetic mean of some sample set X_i that are known to belong to class i . Once the weight vector is determined, the decision to which class the sample vector $x \in X$ belongs can be made by comparing it with the weight vector of each class. The comparison is made through computing the similarity degree based on Yu's norm, which can be written as follows:

$$S\langle x, V \rangle = \max\left(0, (1 + \lambda)\left(S_n(\bar{x}, V) + S_n(x, \bar{V}) - 1 - \lambda S_n(\bar{x}, V) S_n(x, \bar{V})\right)\right), \quad (4)$$

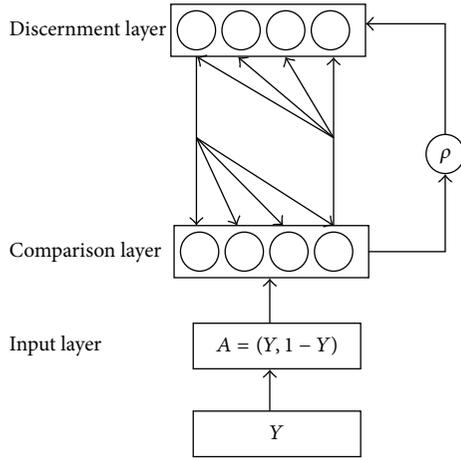


FIGURE 1: The architecture of ART-similarity classifier based on Yu's norm.

where $\bar{x} = 1 - x$ for all $x \in [0, 1]^D$. Also, $\lambda > -1$, affecting the classification accuracy. And as λ is smaller, the classification accuracy is higher generally. Without the loss of generalization here it is set as 0.6. Thus, that sample x which belongs to C_m can be determined by the following:

$$S \langle x, V_m \rangle = \max_{i=1, \dots, N} S \langle x, V_i \rangle. \quad (5)$$

3. ART-Similarity Clustering Method Based on Yu's Norm

The ART neural network and the clustering method using similarity measure based on Yu's norm have their respective advantages as has been noted. The proposed ART-similarity clustering method is the synthesized product of their respective advantages. Its architecture which is shown in Figure 1 is similar to that of the fuzzy ART excluding the adaptive filter that can be adaptive to adjust the number of clusters by the vigilance parameter ρ . It is mainly comprised of input layer storing the input samples, comparison layer receiving bottom-up input from input layer and top-down input from discernment layer, and discernment layer containing the active category and storing category nodes. Each normalized input vector A is denoted as a $2M$ dimensional vector $A(Y, 1 - Y)$, where Y is the normalized data sample and M is the dimensional vector. The weight vector associated with category node in discernment layer is represented as V_j ($j = 1, \dots, N$). Since the proposed clustering method is on the basis of similarity measure, the categories generated are merging of the similar samples. Initially, the weight vector is set to zero vector, and the number of category node is set to zero. For every input sample except the first sample, the categorization with the ART-similarity clustering method is performed by category choice, vigilance test, and learning.

(1) *Category Choice*. The purpose of this stage is to select the winner category node from all existing category nodes. The choice function is a similarity measure between the j th

category and the i th input sample, by (4), which is represented as

$$S \langle Y_i, V_j \rangle = \max \left(0, (1 + \lambda) \left(\text{Sn}(\bar{Y}_i, V_j) + \text{Sn}(Y_i, \bar{V}_j) - 1 - \lambda \text{Sn}(\bar{Y}_i, V_j) \text{Sn}(Y_i, \bar{V}_j) \right) \right), \quad (6)$$

where Y_i and V_j are the i th input sample and the weight vector of the j th category node, respectively. The winner category J which has the biggest similarity degree is selected by the following equation:

$$J = \arg \max_j \left(S \langle Y_i, V_j \rangle \right). \quad (7)$$

(2) *Vigilance Test*. In order to determine whether the i th input sample matches with the selected winner category J in this stage, a vigilance parameter ρ which is introduced as the evaluation criterion of similarity is used to test the similarity degree to which Y_i is a subset of category V_j . If the similarity degree meets the vigilance criterion

$$S \langle Y_i, V_j \rangle \geq \rho, \quad (8)$$

where $\rho \in [0, 1]$, it indicates that the input sample Y_i is sufficiently similar to the selected winner category V_j , the sample Y_i is classified into the J th category, and the learning is also performed. Otherwise, a new category node is generated to contain the input sample in the discernment layer; correspondingly the weight vector of the category is given by the following formula:

$$V_{n+1} = Y_i, \quad (9)$$

where n is total number of current category node.

(3) *Learning*. When the selected category satisfied the vigilance criterion, the weight vector of the current category is updated by the following equation, that is, learning:

$$V'_J = \frac{(n \times V_J + Y_i)}{(n + 1)}, \quad (10)$$

where V'_J is the enhanced weight vector of category J , V_J is the origin weight vector of category J , and n is the number of samples that belong to the category J .

4. Diagnosis System Using ART-Similarity Clustering Method

4.1. *Structure of Diagnosis System*. The fault diagnosis system is shown in Figure 2. The system mainly includes four stages: data acquisition, feature extraction, feature selection, and fault diagnosis. The main objective of the study is to use the proposed ART-similarity clustering method based on Yu's norm to diagnose the different fault categories of the rolling element bearings. In order to ensure the credibility of diagnosis results the vibration signals used are obtained from the dataset of the rolling element bearings [15] in this

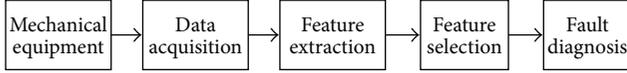


FIGURE 2: Architecture of fault diagnosis system.

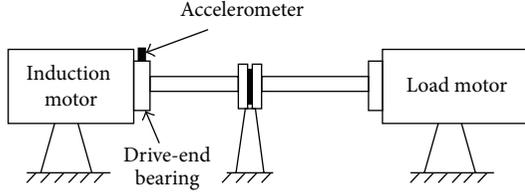


FIGURE 3: The schematic diagram of the experimental setup.

paper. And then time-domain, frequency-domain features, and auto-regression (AR) model parameters are extracted from the raw signals, respectively, and optimized by the modified distance discriminant technique. Finally, the proposed ART-similarity clustering method combined with the salient features is trained and used to diagnose the faults of bearings.

4.2. Vibration Data Acquisition. The schematic diagram of experiment rig is shown in Figure 3. It consists of a three-phase induction motor, a load motor, and a torque sensor. The bearings are installed in a motor driven mechanical system. The load motor is controlled to get the desired torque load levels. An accelerometer is attached to the housing with magnetic bases and mounted at the 12 o'clock position at the driven end of the motor housing. Considering that the frequency content of interest in the vibration signals of the system under study does not exceed 5000 Hz, the vibration signals are acquired by a DAT recorder with the 12 K/s sample rate.

The test bearing type is 6205-2RS JEM SKF, deep groove ball bearing. The single point defects are introduced into the drive-end bearing of the motor by the electrodischarge machining. Four different defect diameters (0.007, 0.014, 0.021, and 0.028 inch) are introduced into the balls to simulate different fault severity of bearings; 0.014 inch defect diameter is introduced into the inner race and outer race, respectively, to simulate the different fault categories of bearings, and these defects' depth is all 0.011 inch; each bearing is tested under four different loads (0, 1, 2, and 3 hp) and rpm \approx 1800. Thus, the bearing data sets can be obtained from the experimental system under different operation loads and seven different fault conditions: (1) in normal condition; (2) with outer race fault; (3) with inner race fault; (4) with four different severity levels of ball faults.

4.3. Features Extraction. Feature parameters are mainly utilized to depict the fault-relate information about the bearings. To acquire more information many different feature parameters are extracted from the vibration signals.

Statistical feature parameters in time domain and frequency domain are often used to characterize the shape of vibration signal from different perspectives. In this study

nine time-domain feature parameters and seven frequency-domain feature parameters are extracted and used as the basis for the fault diagnosis of bearings, which are listed in Table 1.

Time series model can characterize the dynamic process of mechanical system. Because of sensitiveness of these model parameters to the shape of vibration data, these parameters are also used as feature parameters to demonstrate the fault-related information about the bearing. Autoregression (AR) model which is the basis time series model can work as predictor; its basic expression can be written as follows:

$$\bar{s}_n = \varphi_1 s_{n-1} + \varphi_2 s_{n-2} + \cdots + \varphi_r s_{n-r} = \sum_{i=1}^r \varphi_i s_{n-i}, \quad (11)$$

where $s_{n-1}, s_{n-2}, \dots, s_{n-r}$ are the r previous samples, \bar{s}_n is the predicted sample of the signal s_n , and $\varphi_1, \varphi_2, \dots, \varphi_r$ is AR model parameters. The algorithm that these AR model parameters are obtained is described concretely in [16]. Here the parameter r is set as 16.

Thus, an original feature set containing 32 feature parameters is obtained, which can preserve fault-related information that cover time-domain, frequency-domain and wavelet-domain.

4.4. Features Selection. When the above 32 features are used as the input of the proposed ART-similarity clustering method to diagnose the bearings, there is a possibility that the diagnosis accuracy decreases and the computation time is increased because of the redundancy or irrelevance of some features. In order to improve the diagnosis performance, some sensitive features providing characteristic information for the diagnosis system need to be selected, and irrelevant or redundant features must be removed. Here, the distance discriminant technique [17] is adopted to select the salient features from the original feature set. Considering the overlapping degree among different classes, an improved version is proposed.

Assume that a feature set of J classes consists of N samples, and in the j th class there are N_j samples, where $j = 1, 2, \dots, J$, and $N = \sum_{j=1}^J N_j$. Each sample is represented by M features, and the m th feature of the i th sample is written as q_i^m . The feature selection process can be described as follows.

Step 1. Calculate the standard deviation and the mean of all samples in the m th feature:

$$\sigma_m^2 = \frac{1}{N} \sum_{i=1}^N (q_i^m - \bar{q}^m)^2, \quad (12)$$

$$\bar{q}^m = \frac{1}{N} \sum_{i=1}^N q_i^m.$$

Step 2. Calculate the standard deviation and the mean of the sample in the j th class in the m th feature, respectively:

$$\sigma_m^{j2} = \frac{1}{N_j - 1} \sum_{i=1}^{N_j} (q_i^m - \bar{q}_j^m)^2, \quad (13)$$

$$\bar{q}_j^m = \frac{1}{N_j} \sum_{i=1}^{N_j} q_i^m.$$

TABLE 1: Time-domain and frequency-domain feature parameters.

Time-domain feature parameters		Frequency-domain feature parameters	
Feature	Equation	Feature	Equation
Mean (c_1)	$c_1 = \frac{1}{n} \sum_{i=1}^n x_i$	Mean frequency	$p_1 = \sqrt{\frac{\sum_{i=1}^K f_i^2 \cdot S(f_i)}{\sum_{i=1}^K S(f_i)}}$
Root mean square (c_2)	$c_2 = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$	Average frequency that is wave shape of signal crosses the mean of time-domain signal	$p_2 = \sqrt{\frac{\sum_{i=1}^K f_i^4 \cdot S(f_i)}{\sum_{i=1}^K f_i^2 \cdot S(f_i)}}$
Standard deviation (c_3)	$c_3 = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}}$	Stabilization factor of wave shape	$p_3 = \frac{\sum_{i=1}^K f_i^2 \cdot S(f_i)}{\sqrt{\sum_{i=1}^K S(f_i) \sum_{i=1}^K f_i^4 \cdot S(f_i)}}$
Skewness (c_4)	$c_4 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - c_1}{c_2} \right)^3$	Coefficient of variability	$p_4 = \frac{\sigma}{\bar{f}}$
Kurtosis (c_5)	$c_5 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - c_1}{c_2} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$	Frequency-domain skewness	$p_5 = \frac{\sum_{i=1}^K (f_i - \bar{f})^3 \cdot S(f_i)}{\sigma^3 K}$
Crest factor (CF)	$CF = \frac{\max(x)}{c_2}$	Frequency-domain kurtosis	$p_6 = \frac{\sum_{i=1}^K (f_i - \bar{f})^4 \cdot S(f_i)}{\sigma^4 K}$
Latitude factor (LF)	$LF = \frac{\max(x)}{\left((1/n) \sum_{i=1}^n x ^{1/2} \right)^2}$	Root-mean-square ratio	$p_7 = \frac{\sum_{i=1}^K \sqrt{(f_i - \bar{f})} \cdot S(f_i)}{\sqrt{\sigma} \cdot K}$
Shape factor (SF)	$SF = \frac{c_2}{(1/n) \sum_{i=1}^n x_i }$	where $S(f_i)$ is power spectrum density which is obtained by the Welch method. K is the number of spectrum lines used to calculate the parameters, and	
Impulse factor (IF)	$IF = \frac{\max(x)}{(1/n) \sum_{i=1}^n x_i }$	$\bar{f} = \frac{\sum_{i=1}^K f_i \cdot S(f_i)}{\sum_{i=1}^K S(f_i)}, \quad \sigma = \sqrt{\frac{\sum_{i=1}^K (f_i - \bar{f})^2 \cdot S(f_i)}{K}}$	
where x_i is a signal series for $i = 1, 2, \dots, n$, and n is the number of data points.			

Step 3. Calculate the weighted standard deviation of the class center g_j in the m th feature:

$$\sigma_m^{1/2} = \sum_{j=1}^J \rho_j (g_j^m - g^m)^2 = \mu_1 - \mu_2, \quad (14)$$

where $\mu_1 = \sum_{j=1}^J \rho_j (g_j^m)^2$, $\mu_2 = \sum_{j=1}^J \rho_j g_j^m$, $g^m = \sum_{j=1}^J \rho_j g_j^m$, g^m is the center of all samples in the m th feature, g_j^m is the center of the samples of the j th class in the m th feature, μ_1 and μ_2 are the weighted mean of the squared class center g_j^2 and the class center g_j in the m th feature, and ρ_j is the prior probability of the j th class, which can be calculated by the formula $\rho_j = N_j / \sum_{j=1}^J N_j$, $j = 1, 2, \dots, J$ and $\sum_{j=1}^J \rho_j = 1$.

Step 4. Calculate the distance of the m th feature between different classes:

$$d_b^m = \frac{\sigma_m^{1/2}}{\sigma_m^2}. \quad (15)$$

Step 5. Define and calculate the variance factor of d_b^m in the m th feature as follows:

$$u_b^m = \frac{\max \|\bar{q}_i^m - \bar{q}_j^m\|}{\min \|\bar{q}_i^m - \bar{q}_j^m\|}, \quad (16)$$

$$\|\bar{q}_i^m - \bar{q}_j^m\| = \frac{(\bar{q}_i^m - \bar{q}_j^m)}{\sigma_m^2},$$

$i, j = 1, 2, \dots, J, i \neq j.$

Step 6. Calculate the distance of the m th feature within classes:

$$d_w^m = 2 \frac{\sum_{j=1}^J \rho_j \sigma_m^{1/2}(j)}{\sigma_m^2}. \quad (17)$$

Step 7. Define and calculate the variance factor of d_w^m in the m th feature as follows:

$$v_w^m = \frac{\max(d_w^m)}{\min(d_w^m)}. \quad (18)$$

TABLE 2: Statistics of each fault condition of bearing.

Condition	Normal	Inner race fault	Outer race fault	Different severity level of ball fault			
Defect diameter (inch)	0	0.014	0.014	0.007	0.014	0.021	0.028
Label	1	2	3	4	5	6	7
Number of data samples	25	25	25	25	25	25	25

Step 8. The compensation factor of the m th feature can be defined and calculated as follows:

$$\eta_m = \frac{1}{v_w^m} + \frac{1}{u_b^m}. \quad (19)$$

Step 9. Calculate the modified distance discriminant factor of the m th feature:

$$d_b^m - \beta' d_w^m = d_b^m - \eta_m d_w^m = \lambda_m, \quad (m = 1, 2, \dots, M), \quad (20)$$

where β' is used to control the impact of d_w^m .

Step 10. Rank M features in descending order according to the modified distance discriminant factors λ_m . Then normalize λ_m by $\bar{\lambda}_m = (\lambda_m - \min(\lambda_m)) / (\max(\lambda_m) - \min(\lambda_m))$ and get the normalized modified distance discriminant factors. Clearly, bigger $\bar{\lambda}_m$ ($m = 1, 2, \dots, M$) signifies that the corresponding feature is better to separate J classes.

Step 11. Set a threshold value γ , and select the sensitive features whose modified distance discriminant factor $\bar{\lambda}_m \geq \gamma$ from the set of M features.

Further, in order to demonstrate the superiority and character of the improved distance discriminant technique, a numerical example to compare the improved distance discriminant technique with the original distance discriminant technique is presented in the appendix.

4.5. Fault Diagnosis. In the phase of fault diagnosis some data samples of bearings are utilized to evaluate the performance of the proposed method, the data samples contain seven different fault conditions, the fault conditions are labeled by Arabic numerals (1, 2, 3, 4, 5, 6, 7), respectively, and each fault condition contains 25 data samples. The detailed description is shown in Table 2.

The detailed fault diagnosis flow chart of the proposed method is shown in Figure 4. First, data samples are pre-processed to obtain 36 feature parameters. Second, to reduce the computation time and improve diagnosis accuracy the proposed improved distance discriminant technique is used to select the salient features from the original feature set. Figure 5 shows the modified distance discriminant factors $\bar{\lambda}_m$ of all features. From the figure it can be seen that when the threshold $\gamma = 0.88$, the number of selected salient features is 19.

Finally, the proposed ART-similarity clustering method based on Yu's norm is applied to the fault diagnosis of bearings. Its characteristics are training and test together. The 175 data samples are used for training and test. In the

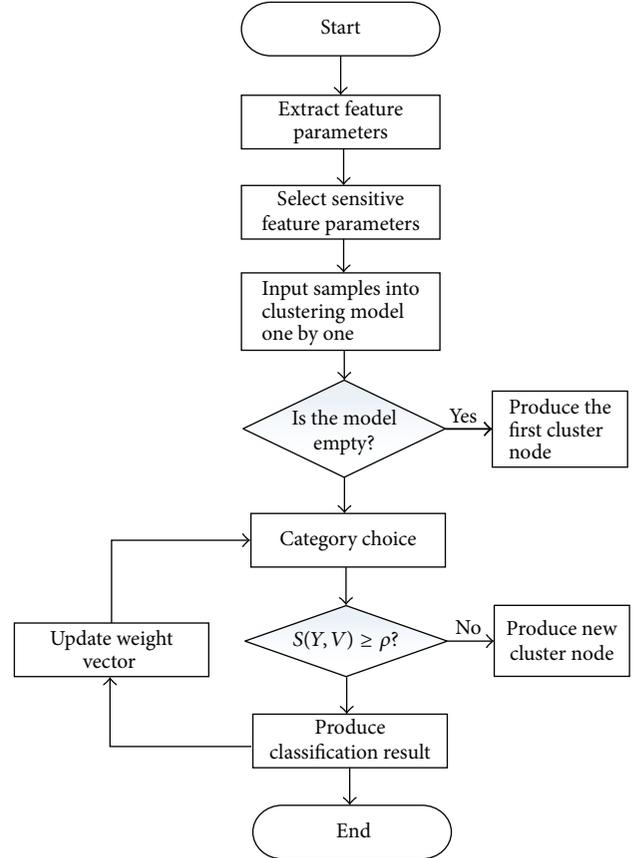


FIGURE 4: Fault diagnosis flowchart of the proposed method.

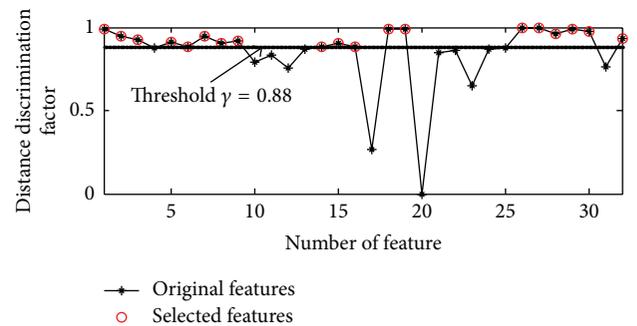


FIGURE 5: Salient feature parameters selection.

beginning of the training of the cluster model, it is empty. When the first data sample is input into the cluster model, the first cluster node is produced which is considered as one fault category. When the next input sample enters the model, it is compared with the first cluster node. If the similarity degree

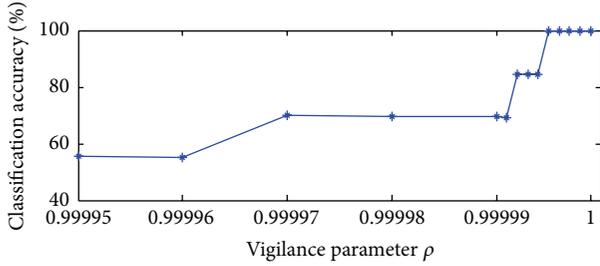


FIGURE 6: Classification accuracy with the different vigilance parameters.

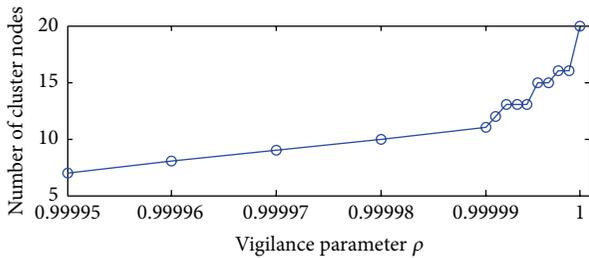


FIGURE 7: Relationship of cluster nodes number and vigilance parameters.

is bigger than the set vigilance parameter ρ , the input sample is classified to the first cluster, and the corresponding weight vector of the cluster node is modified by (7); otherwise, the second cluster node is produced. When the third input sample enters the model, it is compared with all the produced cluster nodes, and the cluster node that has the biggest degree of similarity is the winner. If the similarity degree meets the vigilance criterion, the sample is classified to the winner cluster, or else a new cluster node is produced. According to the above-mentioned reasoning, a trained classifier is obtained.

Generally, one fault class is needed to use many cluster nodes to learn because of the complex fault mechanism. To evaluate the performance of the proposed Yu's norm based on ART-similarity clustering method and understand the relationship of classification accuracy, the number of cluster nodes, and vigilance parameter, a series of fault diagnosis experiments with the increasing vigilance parameters are conducted. For convenience of computation, the classification accuracy can be obtained by the following formula [8]:

$$DA = \frac{C}{(T - N)} \times 100\%, \quad (21)$$

where DA is classification accuracy, C is the sample number of correct classification, T is the number of total samples, and N is the number of generated cluster nodes. $T - N$ which is the difference of the total sample number minus the training sample number means the number of used samples for test.

Figure 6 shows the relationship of classification accuracy and the increasing vigilance parameters. It can be seen that the classification accuracy rises with the increasing vigilance parameter ρ from the figure, but it is not continuous. As $\rho \geq 0.999995$ the classification accuracy reaches 100%. Figure 7

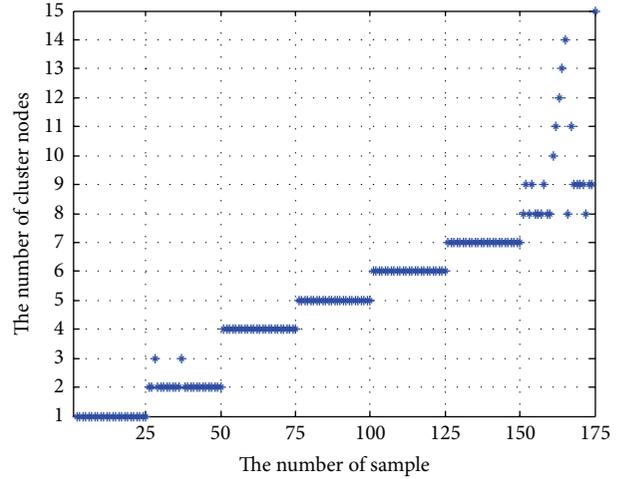


FIGURE 8: Classification of all data samples.

TABLE 3: Number of neurons presenting each condition.

Label	1	2	3	4	5	6	7	Total
Number of neurons	1	2	1	1	1	1	8	15

shows the relationship of cluster nodes number and the increasing vigilance parameters. The number of cluster nodes rises with the increasing vigilance parameters and each fault class can be composed of many cluster nodes. When $\rho = 0.999999$, the number of cluster nodes which covers 7 fault conditions of bearings is as high as 20. And from these two figures it can be seen that when $\rho < 0.99999$, the number of cluster nodes is about 10, but the classification accuracy is very low and is lower than 75%; when $\rho > 0.99999$ and rises, the classification accuracy and the number of cluster nodes all increase; when $\rho > 0.999995$, the classification accuracy is the highest and reaches 100%, and the corresponding number of cluster nodes is 15 and is the least relatively. Thus according to the relationship of cluster nodes number and vigilance parameter ρ and classification accuracy, ρ is set as 0.999995 here.

For convenience of understanding, Figure 8 and Table 3 show the classification result using all samples for the training and test when $\rho = 0.999995$. From the table and figure it can be seen that all samples are classified accurately, and each condition of bearing includes different cluster nodes number. Conditions 1, 3, 4, 5, and 6 only use one node to learn and test, respectively, but the condition 7, namely, very severe bearing ball fault, needs eight cluster nodes. This is mainly because the region of the cluster node becomes bigger or smaller depending on the space distribution of the data samples with the same condition. When the node region becomes small, the corresponding number of cluster nodes increases.

4.6. Performance Test of ART-Similarity Clustering Method Based on Bootstrap Method. It is well known that the initial conditions affect the performance of ART-similarity clustering method. To study the stability and generalization of the

TABLE 4: Statistical performance of ART-similarity clustering method.

Clustering method	Mean (%)	Standard deviation (%)	Confidence interval (%)
ART-similarity	98.95	0.54	[97.87, 99.58]

TABLE 5: Comparison of classification with different neural networks.

Classification method	Fuzzy ART	SOFM	ART-similarity
Number of cluster nodes	79	76	15
Classification accuracy (%)	96.57	94.36	100

proposed method, the bootstrap method is used to compute the estimated mean, standard deviation, and confidence interval for the classification accuracy, which is useful for estimating a parameter when the underlying distribution function of parameter is unknown [18]. 1000 bootstrap replicate samples are generated for statistical analysis of diagnosis accuracy. The estimated statistical results are given in Table 4. From the table it can be observed that the estimated statistical mean of diagnosis result is 98.95%, the standard deviation is 0.54%, and the 95% confidence interval achieves [97.87%, 99.38%]. These all can indicate that the performance of the proposed ART-similarity clustering method is stable and generalized.

4.7. Classification Performance Comparison with Other Methods. In order to validate the superiority of the proposed ART-similarity clustering method, the classification result produced by ART-similarity classifier is compared with that produced by other conventional unsupervised neural networks, such as the fuzzy ART and SOFM network. Same data samples are utilized to evaluate these methods: 100 of these samples are used for the training of these networks, and the rest, for test. The classification results of the ART-similarity clustering method versus other classification methods with the same salient feature parameters are shown in Table 5. From the table it can be drawn that the classification accuracy of the ART-similarity is the highest and reaches 100%, and the corresponding number of cluster nodes is the least and is 15. For the fuzzy ART and SOFM, the classification accuracy is 96.57% and 94.36%, and the corresponding number of cluster nodes is 79 and 76, respectively. These indicate that the proposed ART-similarity clustering method has superior classification performance comparatively.

5. Conclusions

In this paper a new clustering method that combines the adaptive resonance theory (ART) with the similarity measure based on Yu's norm is presented to diagnose the faults of rolling element bearings, which can generate the cluster nodes dynamically. Before application of the proposed clustering method to the fault diagnosis of bearings, time-domain statistical characteristics features, frequency-domain statistical characteristics features, and AR time series model param-

eters are extracted to characterize the fault-related information of bearing.

Owing to the redundancy and irrelevance of some features the improved distance discriminant techniques are used to select the sensitive features, and then they are input into the proposed clustering method to diagnose the fault categories of bearings. The experiment result showed that the proposed ART-similarity clustering method can diagnose the faults of bearings successfully, and its diagnosis accuracy is higher than fuzzy ART and SOFM. And because the initial conditions affect the performance of the proposed clustering method, the bootstrap method is utilized to analyze the diagnosis results. The statistical analysis result shows that the proposed clustering method is stable and generalized. All these indicate that the proposed method has better diagnosis ability and performance and further demonstrate that the proposed clustering method has a good promise in the field of fault diagnosis of mechanical system.

Appendix

Comparison of the Modified Distance Discriminant Technique with the Original Distance Discriminant Technique: A Numerical Example

Suppose the simulated data set is comprised of four classes. Each class consists of 35 samples, and each sample is depicted by six features. Figure 9 shows the distributions of six features under four classes. From the figure, it can be seen clearly that the classification ability of each feature is listed as follows: $ax_1 > ax_2 > ax_3 > ax_4 > ax_5 > ax_6$, where ax_i ($i = 1, 2, \dots, 6$) represent the indexes of the classification abilities of the six features, respectively.

To ascertain classification ability of each feature, the distance discriminant technique is adopted to evaluate the sensitivities of the six features. The evaluation results are shown in Figure 10(a), and the modified evaluation results produced by the modified distance discriminant technique are displayed in Figure 10(b). By comparison of Figure 10(a) with Figure 10(b), it can be found that the original distance discriminant factors of the six features produced by the modified distance discriminant technique are in accord with the truth. But, the distance discriminant factor produced by the distance discriminant technique of feature x_4 is bigger than that of feature x_3 , and factor of feature x_6 is bigger than that of feature x_5 ; namely, the sensitivities of feature x_4 and x_6 are higher than that of feature x_3 and x_5 , respectively. This is because the distance between different classes and distance within classes are obtained by the simple mean method in the original distance discriminant technique, and the variance factors of the two distances are not considered. However, in the modified distance discriminant technique, the variance factors are incorporated into the distance between different classes and distance within classes. Thus, the drawbacks of the original distance discriminant technique are overcome, and the modified distance discriminant factors shown in Figure 10(b) are produced. These demonstrate further that the

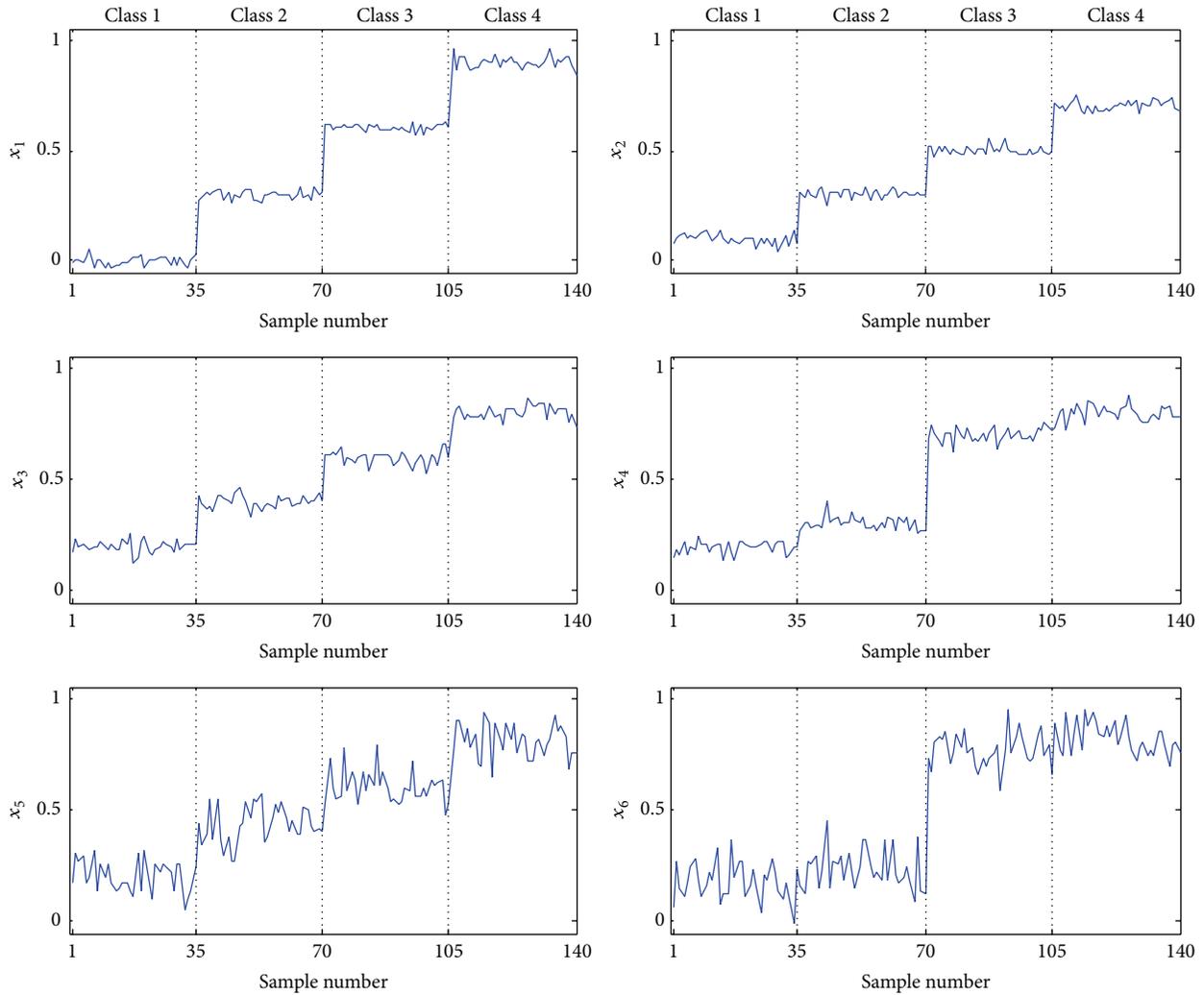


FIGURE 9: The distribution of six features under four classes, where x_i ($i = 1, 2, \dots, 6$) represent features.

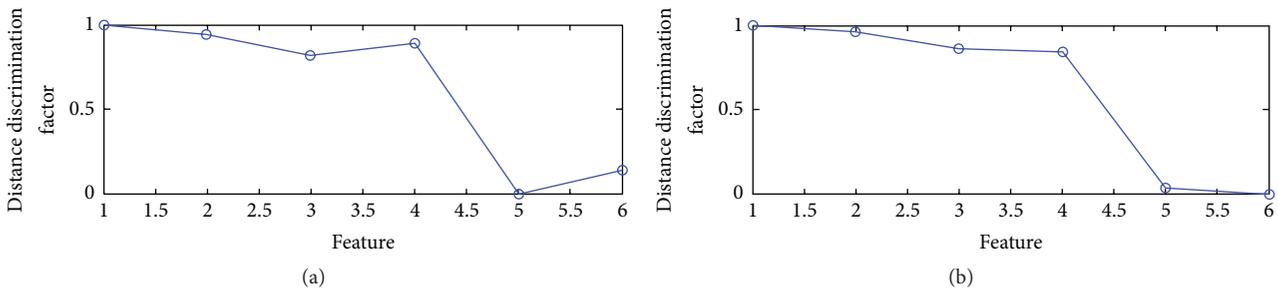


FIGURE 10: The distance discriminant factors of the six features (a) produced by the distance discriminant technique and (b) produced by the modified distance discriminant technique.

modified distance discriminant technique is superior to the original distance discriminant technique.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The work is supported financially by the National Natural Science Foundation of China (Grant No. 51405353), and the Open Fund of the Key Laboratory for Metallurgical Equipment and Control of Ministry of Education in Wuhan University of Science and Technology (2014B01), and National

Natural Science Foundation of China (Grant nos. 51475339 and 51575202).

References

- [1] M. M. Polycarpou and A. J. Helmicki, "Automated fault detection and accommodation: a learning system approach," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, no. 11, pp. 1447–1458, 1995.
- [2] S. Grossberg, "Adaptive resonance theory: how a brain learns to consciously attend, learn, and recognize a changing world," *Neural Networks*, vol. 37, pp. 1–47, 2013.
- [3] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 698–713, 1992.
- [4] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, no. 6, pp. 759–771, 1991.
- [5] G. A. Carpenter and S. Grossberg, "ART 2: self-organization of stable category recognition codes for analog input patterns," *Applied Optics*, vol. 26, no. 30, pp. 4919–4930, 1987.
- [6] I. Dagher, M. Georgiopoulos, G. L. Heileman, and G. Bebis, "An ordering algorithm for pattern presentation in fuzzy ARTMAP that tends to improve generalization performance," *IEEE Transactions on Neural Networks*, vol. 10, no. 4, pp. 768–778, 1999.
- [7] T.-C. Lin and P.-T. Yu, "Centroid neural network adaptive resonance theory for vector quantization," *Signal Processing*, vol. 83, no. 3, pp. 649–654, 2003.
- [8] B. S. Yang, T. Han, and J. L. An, "ART-KOHONEN neural network for fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 18, no. 3, pp. 645–657, 2004.
- [9] T. Wakahara and Y. Yamashita, " k -NN classification of handwritten characters via accelerated GAT correlation," *Pattern Recognition*, vol. 47, no. 3, pp. 994–1001, 2014.
- [10] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4137–4145, 2013.
- [11] L. Szilágyi and S. M. Szilágyi, "Generalization rules for the suppressed fuzzy c -means clustering algorithm," *Neurocomputing*, vol. 139, pp. 298–309, 2014.
- [12] P. Luukka, "Similarity classifier using similarity measure derived from Yu's norms in classification of medical data sets," *Computers in Biology and Medicine*, vol. 37, no. 8, pp. 1133–1140, 2007.
- [13] L. Silva, R. Moura, A. M. P. Canuto, R. H. N. Santiago, and B. Bedregal, "An interval-based framework for fuzzy clustering applications," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 2174–2187, 2015.
- [14] H. J. Zimmermann, *Fuzzy Set Theory—And Its Applications*, Kluwer Academic, New York, NY, USA, 2001.
- [15] K. A. Loparo, *Bearings Vibration Data Set*, Case Western Reserve University, Cleveland, Ohio, USA, 2003.
- [16] Y. Shuzi, W. Ya, and X. Jianping, *Time Series Analysis in Engineering Application*, Huazhong University of Science and Technology Press, 2007.
- [17] J. N. Liang, S. Yang, and A. Winstanley, "Invariant optimal feature selection: a distance discriminant and feature ranking based solution," *Pattern Recognition*, vol. 41, no. 5, pp. 1429–1439, 2008.
- [18] B. Efron, "Bootstrap methods: another look at the jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

