Clustering algorithm–based fault diagnosis

Chapter · December 2017

DOI: 10.1016/B978-0-12-811534-3.00004-4

1 author:

Yaguo Lei
Xi'an Jiaotong University

Available online at www.sciencedirect.com

ScienceDirect

Mechanical Systems and Signal Processing 22 (2008) 419–435

Mechanical Systems and Signal Processing

www.elsevier.com/locate/jnlabr/ymssp

New clustering algorithm-based fault diagnosis using compensation distance evaluation technique

Yaguo Lei[a,*], Zhengjia He[a,b], Yanyang Zi[a], Xuefeng Chen[a]

[a]School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China
[b]State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China

Received 1 April 2007; received in revised form 25 July 2007; accepted 27 July 2007
Available online 6 August 2007

Abstract

This paper presents a fault diagnosis method of rotating machinery based on a new clustering algorithm using a compensation distance evaluation technique (CDET). A two-stage feature selection and weighting technique is adopted in this algorithm. Feature weights are computed via CDET according to the sensitivity of features and assigned to the corresponding features to indicate their different importance in clustering. Feature weighting highlights the importance of sensitive features and simultaneously weakens the interference of insensitive features. The new clustering algorithm is described and applied to incipient fault and compound fault diagnosis of locomotive roller bearings. The diagnosis result shows the algorithm is able to reliably recognise not only different fault categories and severities but also the compound faults, and demonstrates the superior effectiveness and practicability of the algorithm. Therefore, it is a promising approach to fault diagnosis of rotating machinery.

Keywords: Feature weighting; Compensation distance evaluation technique; Clustering algorithm; Fault diagnosis

1. Introduction

With the rapid development of scientific technology, mechanical equipment in modern industry is growing larger, more precise and more automatic. Their structures become more complex and their potential faults become more difficult to find. Rotating machinery covers a broad range of mechanical equipment and plays an important role in industrial applications. Typical applications are in aeronautical, naval and automotive industries. The need to decrease the downtime on production machinery and to increase reliability against possible failures has attracted interest in fault diagnosis of these systems in recent years. The main purpose of fault diagnosis is to analyse the relevant external information in order to judge the condition of the inaccessible internal components so as to decide if the machine needs to be dismantled or not [1–3]. One of the principal tools for diagnosing rotating machinery problems is the vibration analysis [4–6]. Through the use of some processing techniques of vibration signals, it is possible to obtain vital diagnosis information from the

*Corresponding author. Tel.: +86 29 82667963.
E-mail address: leiyaguo@163.com (Y. Lei).

doi:10.1016/j.ymssp.2007.07.013

vibration signals. However, many techniques presently available require a good deal of expertise to apply them successfully. Simpler approaches are needed which allow relatively unskilled operators to make reliable decisions without the need for a diagnosis specialist to examine data and diagnose problems. Therefore, there is a demand for techniques that can make decision on the running health of the machine automatically and reliably [7–9].

Fuzzy clustering, owing to its superiority in dealing with uncertainty and independent of supervisors, has been widely studied and applied to automatic fault diagnosis [10,11]. According to the principle that "similar objects are within the same cluster and dissimilar objects are in different clusters", a fuzzy clustering algorithm employs a fuzzy mathematics method to partition a data set into several homogenous groups or clusters. During clustering, there is no teacher to provide guidance; hence, it is also called unsupervised classification. The main characteristic of the fuzzy clustering algorithm is that each sample is subject to one cluster with a certain grade of membership [12,13].

Among various clustering algorithms, the fuzzy *c*-means (FCM) algorithm [14,15] is one of the most well-known and widely used algorithms [16]. However, the FCM algorithm supposes that all features have the uniform contributions to clustering, and does not take account of their different importance degrees. Therefore, the FCM algorithm frequently results in inexact clustering results.

There are several improved clustering algorithms reported in the recent literature, in which the importance of the features in clustering is considered. Elaine et al. [17] developed an attribute-weighting clustering algorithm which is achieved by the development of a new procedure to generate the weight for each attribute. Wang et al. [18] introduced a weighted FCM algorithm based on weighted Euclidean distance. Hichem and Olfa [19] presented a new approach to perform clustering and feature weighting simultaneously, and used it to segment colour images.

Generally, some features of a feature set are irrelevant or redundant in fault diagnosis of rotating machinery. If the whole feature set is employed to fault diagnosis directly, it will make the diagnosis process slower and the diagnosis accuracy lower. Thus, to improve the diagnosis accuracy and reduce the computation, a few features which obviously characterise the machine conditions need to be selected from the original feature set. Due to the simpleness and reliability of the distance evaluation technique, it is generally adopted in fault diagnosis [20–23]. Here, a two-stage feature selection and weighting technique based on a compensation distance evaluation technique (CDET) is developed and applied to feature selection and feature weighting.

Based on the above principles, an improvement to the FCM algorithm is made and a new clustering algorithm to fault diagnosis is proposed. The proposed algorithm is created by incorporating the two-stage feature selection and weighting based on CDET into the FCM algorithm. Feature weights are computed via CDET and assigned to the corresponding features to reflect their different sensitivities. In comparison with the existing algorithms [17–19], the proposed algorithm harnesses the merits that the computation of feature weights is simpler and the weights are easier to be understood. The algorithm is applied to fault diagnosis of locomotive roller bearings. The vibration signals are measured from the roller bearings under different conditions including the different fault categories, fault severities and compound faults. The diagnosis result validates the effectiveness and practicability of the new clustering algorithm based on CDET.

2. Experimental system

Roller bearings are at the heart of rotating machinery. In rotating machinery, the failure of roller bearings can result in the deterioration of machine running conditions. Therefore, it is significant to be able to accurately and automatically detect and diagnose the existence and severity of the faults occurring in the bearings.

The test bench of a locomotive roller bearing is shown in Fig. 1. The test bench consists of a hydraulic motor, two supporting pillow blocks (mounting with normal bearing), a test bearing (52732QT) which is loaded on the outer race by a hydraulic cylinder, a hydraulic radial load application system, and a tachometer for shaft speed measurement. The bearing is installed in a hydraulic motor driven mechanical system. 608A11-type ICP accelerometers with a bandwidth up to 5000 Hz are mounted on the load module adjacent to the

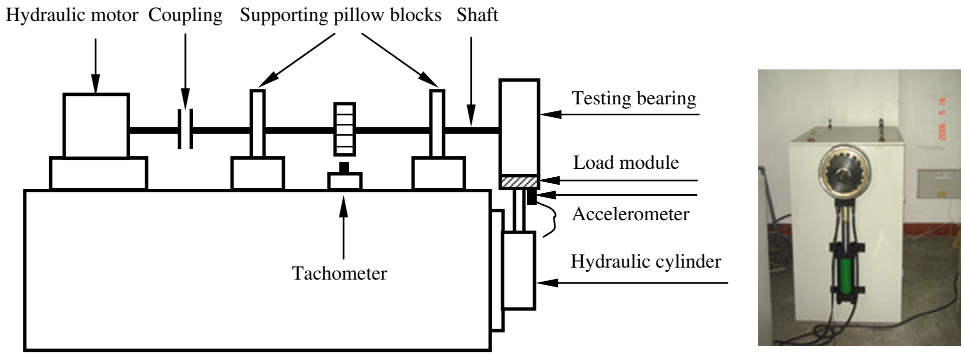

Fig. 1. Test bench of the locomotive roller bearing.

Table 1
Parameters in the experiment

Parameter	Value
Bearing specs	552732QT
Load	9800 N
Inner race diameter	160 mm
Outer race diameter	290 mm
Roller diameter	34 mm
Roller number	17
Contact angle	0°
Sampling frequency	12.8 kHz

outer race of the test bearing for measuring its vibrations. The Advanced Data Acquisition and Analysis System by Sony EX is used to collect the data. Some parameters in the experiment are listed in Table 1.

A bearing data set containing nine subsets is obtained from the experimental system under the nine different operating conditions. The nine conditions of faulty bearings (include normal condition) are described in Table 2. In the nine conditions of bearing, serious fault in the outer race (condition 2) is serious flaking fault, and the other faults are all slight rub. These faulty bearings are shown in Fig. 2. Each data subset corresponds to one of the nine conditions and it consists of 50 samples. Each sample is a vibration signal containing 8192 sampling points. Figs. 3(a)–(i) give a typical raw data sample for each condition, its spectrum and Hilbert envelope spectrum for each condition. Figs. 3(a)–(i), respectively, show the nine conditions of bearing: normal condition, slight rub fault in the outer race, serious flaking fault in the outer race, slight rub fault in the inner race, roller rub fault, compound faults in the outer and inner races, compound faults in the outer race and rollers, compound faults in the inner race and rollers and compound faults in the outer and inner races and rollers. The rotating speed is listed in Table 2, respectively, when each data was collected. From the raw vibration signals and the corresponding FFT spectrums, it is not easy to identify the different faults. With the envelope spectrums, some faults, such as slight rub of outer race, serious flaking of outer race, and roller rub faults may be recognised. However, it is extremely difficult to classify the nine different conditions only using the raw vibration signals, FFT, and envelope spectrums. Thus, in order to identify the nine different conditions of bearing accurately, it is necessary to apply a more efficient method to extract the fault features.

3. Feature extraction and weighting

3.1. Feature extraction

In this study, both time-domain and frequency-domain feature parameters, which are listed in Table 3, are respectively extracted to recognise various operation conditions of rotating machinery.

Table 2
Description of the faulty bearings

Condition	Rotating speed (rpm)	Label
Normal condition	About 490	1
Slight rub fault in the outer race	About 490	2
Serious flaking fault in the outer race	About 480	3
Slight rub fault in the inner race	About 500	4
Roller rub fault	About 530	5
Compound faults in the outer and inner races	About 520	6
Compound faults in the outer race and rollers	About 520	7
Compound faults in the inner race and rollers	About 640	8
Compound faults in the outer and inner races and rollers	About 550	9

Fig. 2. Faults in the locomotive roller bearings.

When faults occur in rotating machinery, the time-domain signal may change. Both its amplitude and distribution may be different from those of the time-domain signal under normal condition. Also, the frequency spectrum of the time-domain signal and the distribution of the frequency spectrum may change, which signifies that new frequency components may appear and a change of the convergence of the frequency spectrum may take place. x_m, x_{rms} and x_p may reflect the vibration amplitude and energy in time domain. x_{std}, x_{ske}, x_{kur}, CF, CLF, SF and IF may represent the time series distribution of the signal in time domain. x_{mf} may

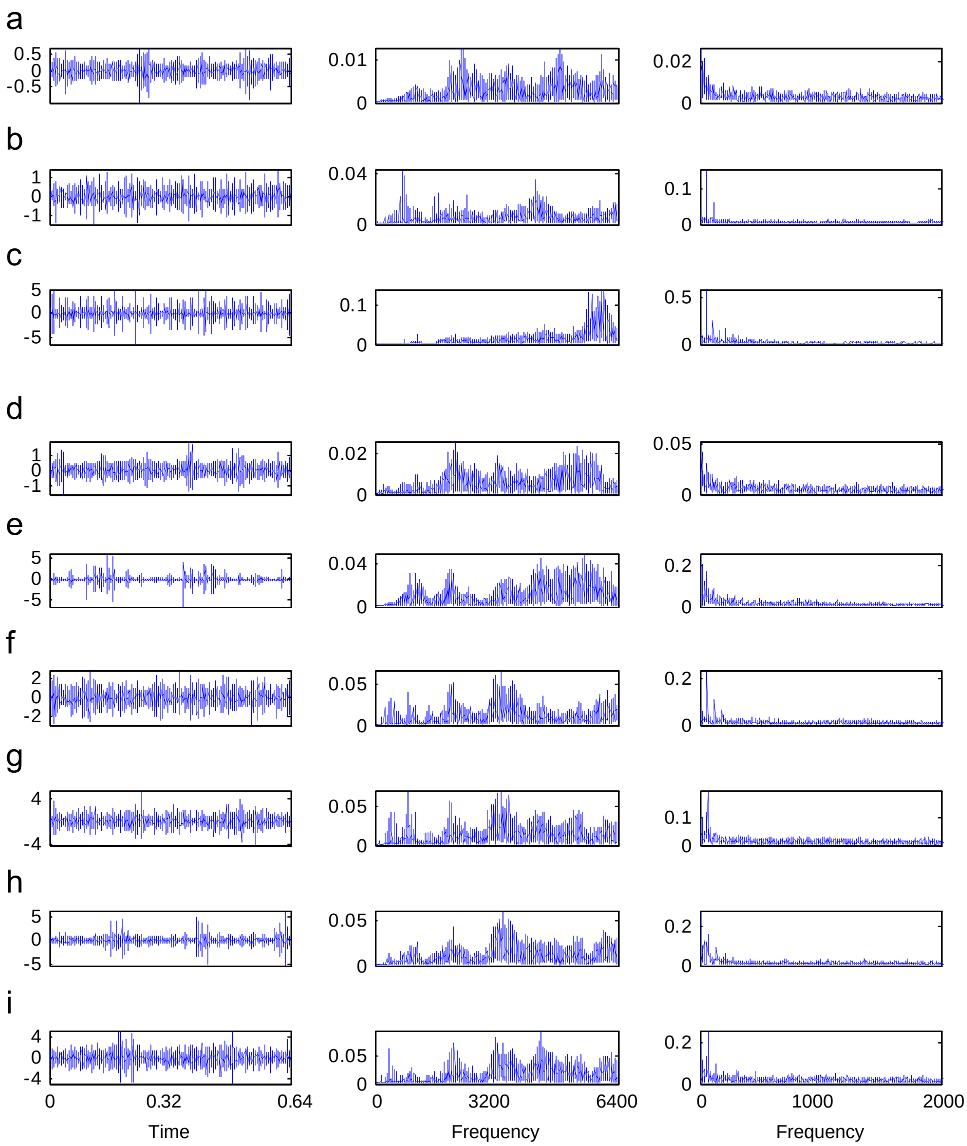

Fig. 3. Typical raw data sample, its spectrum, and envelope spectrum for nine conditions.

indicate the vibration energy in frequency domain. x_{fc} and x_{rmsf} may show the position change of main frequencies. x_{stdf} may describe the convergence of the spectrum power. SPRO, SPRI and SPRR may reflect the defect occurrence in the outer race, inner race, and roller of bearings, respectively.

Some of the above feature parameters have been demonstrated not effective in previous publications. But in different paper, different feature parameters are applied according to the experience accumulated by different researchers. In the different applications, different feature parameters behave as different diagnosis performance. One of the purposes of this paper is to select feature parameters automatically using a method less dependent on human experience. Thus, many feature parameters are calculated in the study. Aiming at different cases, sensitive features may always be selected with the proposed distance evaluation technique.

First, 10 time-domain features are extracted from each of the vibration signals collected, and four frequency-domain features x_{mf}, x_{fc}, x_{rmsf} and x_{stdf} are also extracted from its FFT spectrum. Therefore, 14 feature values are obtained.

Table 3
The feature parameters

Time-domain feature parameters		Frequency-domain feature parameters	
Feature	Equation	Feature	Equation
Mean	$x_m = \sum_{n=1}^{N} x(n) N$	Mean frequency	$x_{mf} = \frac{\sum_{k=1}^{K} s(k)}{K}$
Standard deviation	$x_{\text{std}} = \sqrt{\frac{\sum_{n=1}^{N}(x(n)-x_m)^2}{N-1}}$	Frequency centre	$x_{fc} = \frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$
Root mean square	$x_{\text{rms}} = \sqrt{\frac{\sum_{n=1}^{N}(x(n))^2}{N}}$	Root mean square frequency	$x_{\text{rmsf}} = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)}}$
Peak	$x_p = \max\lvert x(n)\rvert$	Standard deviation frequency	$x_{\text{stdf}} = \sqrt{\frac{\sum_{k=1}^{K}(f_k - x_{fc})^2 s(k)}{\sum_{k=1}^{K} s(k)}}$
Skewness	$x_{\text{ske}} = \frac{\sum_{n=1}^{N}(x(n)-x_m)^3}{(N-1)x_{\text{std}}^3}$	Spectrum peak ratio outer	$\text{SPRO} = \frac{K\sum_{i=1}^{H} p_{\text{O}}(h)}{\sum_{k=1}^{K} s(k)}$
Kurtosis	$x_{\text{kur}} = \frac{\sum_{n=1}^{N}(x(n)-x_m)^4}{(N-1)x_{\text{std}}^4}$	Spectrum peak ratio inner	$\text{SPRI} = \frac{K\sum_{i=1}^{H} p_{\text{I}}(h)}{\sum_{k=1}^{K} s(k)}$
Crest factor	$CF = \frac{x_p}{x_{\text{rms}}}$	Spectrum peak ratio roller	$\text{SPRR} = \frac{K\sum_{i=1}^{H} p_{\text{R}}(h)}{\sum_{k=1}^{K} s(k)}$
Clearance factor	$\text{CLF} = \frac{x_p}{(\frac{1}{N}\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert})^2}$		
Shape factor	$SF = \frac{x_{\text{rms}}}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$		
Impulse factor	$IF = \frac{x_p}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$		

Time-domain: where $x(n)$ is a signal series for $n = 1, 2, \ldots, N$, N is the number of data points.

Frequency-domain: where $s(k)$ is a spectrum for $k = 1, 2, \ldots, K$, K is the number of spectrum lines; f_k is the frequency value of the kth spectrum line; $p_{\text{O}}(h)$, $p_{\text{I}}(h)$ and $p_{\text{R}}(h)$ are, respectively, the peak values of the hth ($h = 1, 2, \ldots, H$, H is the number of harmonics) harmonics of the characteristic frequencies for bearing outer race (f_{O}), inner race (f_{I}) and roller (f_{R}), which can be calculated according to the following equations:

$$f_{\text{O}} = \frac{f_r}{2} N_{\text{R}}\left(1 - \frac{B}{C}\cos\alpha\right), \quad f_{\text{I}} = \frac{f_r}{2} N_{\text{R}}\left(1 + \frac{B}{C}\cos\alpha\right) \quad \text{and} \quad f_{\text{R}} = \frac{f_r C}{2B}\left[1 - \left(\frac{B}{C}\cos\alpha\right)^2\right],$$

f_r is the shaft rotational frequency; N_{R} is the roller number; α is the contact angle; B and C are the roller and pitch diameters, respectively.

Second, the examination of the vibration signals acquired indicates the presence of low-frequency interference. The signals are subjected to either high-pass or band-pass filtration to remove the low-frequency interference components. One band-pass and one high-pass filters are adopted. The band-pass frequency of the band-pass filter is chosen as 1.5–4.0 kHz. The cut-off frequency of the high-pass filter is chosen as 3.0 kHz. These frequencies are selected to cover the signal components containing the majority of the roller bearing energy. The 10 time-domain features shown in Table 3 are extracted from each of these filtered signals. Thus, 10×2 time-domain features are produced.

Finally, the effects of interfering signals within the selected frequency band can be minimised by demodulation. Demodulation detection makes the diagnosis process a little more independent of a particular machine since it focuses on the low-amplitude high-frequency broadband signals characterising bearing conditions [24,25]. The Hilbert demodulation is performed on the band-pass and high-pass filtered signals, respectively. Each Hilbert envelope spectrum of the filtered signals is further processed to extract the seven frequency-domain features and another 7×2 frequency-domain feature values are acquired (when SPRO, SPRI and SPRR are extracted, the number H of harmonics is selected as 3). Thus, altogether 48 feature values are obtained.

3.2. A two-stage feature selection and weighting technique based on CDET

It is obvious that different features have different importance degrees to identify the different faults. Some features are sensitive and closely related to the faults, but others are not. Thus, in order to improve the diagnosis performance and avoid the curse of dimensionality, sensitive features providing bearing fault-related information need be selected and highlighted, and irrelevant or redundant features must be discarded or weakened. Here, a two-stage feature selection and weighting technique is presented, namely, which consists of two stages: feature selection and feature weighting.

3.2.1. Stage 1: feature selection

Feature selection is performed with CDET, which is the improved version of distance evaluation technique employed in Refs. [20–23]. It can be described as follows.

Suppose that a feature set of C conditions is

$$\{q_{m,c,j}, m = 1,2,\ldots,M_c; c = 1,2,\ldots,C; j = 1,2,\ldots,J\}, \tag{1}$$

where $q_{m,c,j}$ is the jth feature value of the mth sample under the cth condition, M_c is the sample number of the cth condition, and J is the feature number of each sample. M_c samples under the cth condition may be collected. Therefore, $M_c \times C$ samples are obtained for C classes. For each sample, J features may be extracted to represent the sample. Thus, $M_c \times C \times J$ features may be obtained. And these features are defined as the feature set $\{q_{m,c,j}\}$.

Then the feature selection process based on CDET can be summarised in Table 4.

Table 4
The process of feature selection

(1) Calculating the average distance of the same condition samples

$$d_{c,j} = \frac{1}{M_c \times (M_c - 1)} \sum_{l,m=1}^{M_c} |q_{m,c,j} - q_{l,c,j}|, \quad l,m = 1,2,\ldots,M_c, l \neq m;$$

then getting the average distance of C conditions $d_j^{(w)} = (1/C)\sum_{c=1}^{C} d_{c,j}$.

(2) Defining and calculating the variance factor of $d_j^{(w)}$as follows: $v_j^{(w)} = \max(d_{c,j})/\min(d_{c,j})$.

(3) Calculating the average feature value of all samples under the same condition $u_{c,j} = (1/M_c)\sum_{m=1}^{M_c} q_{m,c,j}$; then obtaining the average distance between different condition samples

$$d_j^{(b)} = \frac{1}{C \times (C-1)} \sum_{c,e=1}^{C} |u_{e,j} - u_{c,j}|, \quad c,e = 1,2,\ldots,C, c \neq e.$$

(4) Defining and calculating the variance factor of $d_j^{(b)}$ as follows:

$$v_j^{(b)} = \frac{\max(|u_{e,j} - u_{c,j}|)}{\min(|u_{e,j} - u_{c,j}|)}, \quad c,e = 1,2,\ldots,C, c \neq e.$$

(5) Defining and calculating the compensation factor as follows:

$$\lambda_j = \frac{1}{\frac{v_j^{(w)}}{\max(v_j^{(w)})} + \frac{v_j^{(b)}}{\max(v_j^{(b)})}}.$$

(6) Calculating the ratio $d_j^{(b)}$ and $d_j^{(w)}$ and assigning the compensation factor $\alpha_j = \lambda_j(d_j^{(b)}/d_j^{(w)})$; then normalising α_j by its maximum value and getting the distance evaluation criteria $\bar{\alpha}_j = \alpha_j/\max(\alpha_j)$.

(7) Setting a threshold value $\phi(\phi \in [0,1])$ and selecting the sensitive features whose distance evaluation criteria $\bar{\alpha}_j \geqslant \phi$ from the feature set $q_{m,c,j}$.

It is clear that bigger $\overline{\alpha}_j$ ($j = 1, 2, \ldots, J$) signifies that the corresponding feature is better to separate the C conditions. Therefore, the sensitive features may be selected from the feature set when their distance evaluation criteria $\overline{\alpha}_j \geqslant \phi$.

3.2.2. Stage 2: feature weighting

Although the sensitive features have been selected from the original feature set via CDET, the selected features have various sensitivities in fault diagnosis. Thus, feature weighting is absolutely necessary to achieve a more reliable diagnosis result. Feature weighting can be regarded as a generalisation of feature selection, and it is that a number within [0, 1] is assigned to a feature for indicating the sensitivity of this feature. In the Euclidean space, feature weighting is to extend the axes corresponding to the sensitive features and shrink the axes corresponding to the features unrelated to fault.

Firstly, a few features are attained from the original feature after feature selection of stage 1. Secondly, CDET is performed again to evaluate the selected features. Then, the final distance evaluation criteria of the selected features may be obtained. Finally, the final distance evaluation criteria are acted as feature weights **wf** ($\mathbf{wf} = (wf_1, \ldots, wf_m, \ldots, wf_M)$, M is the number of the selected features) and assigned to each of the selected features to point out their sensitivities in fault diagnosis.

3.3. Normalisation

The importance of normalisation to the success of clustering algorithms has been proved. Prior to clustering, all features have been normalised. The feature value v_t of the tth feature parameter is normalised by the following formula:

$$v'_t = \frac{v_t - \min(v_t)}{\max(v_t) - \min(v_t)}, \quad t = 1, 2, \ldots, T, \tag{2}$$

where T is the feature number.

4. The new clustering algorithm

The new clustering algorithm proposed in this paper is created by combining the FCM algorithm with the two-stage feature selection and weighting technique presented in the preceding section. It is similar to the FCM algorithm, and performs clustering by minimising the following objective function. The flow chart of the proposed algorithm is shown in Fig. 4.

The objective function J of the new clustering algorithm is expressed as:

$$J(\mathbf{U}, \mathbf{Z}, \mathbf{wf}; \mathbf{X}) = \sum_{k=1}^{K} \sum_{i=1}^{N} (u_{ik})^{\lambda} (d_{ik}^{(wf)})^2 \tag{3}$$

subject to

$$\sum_{k=1}^{K} u_{ik} = 1, \quad 0 \leqslant u_{ik} \leqslant 1, \tag{4}$$

where $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N\}$ is a data set, $\mathbf{x}_i = (x_{i1}, \ldots, x_{im}, \ldots, x_{iM})$ is the ith sample for $i = 1, \ldots, N$ and $m = 1, \ldots, M$, N is the number of samples, M is the number of the features, and x_{im} is its mth feature. $\mathbf{U} = [u_{ik}]$ is the fuzzy partition matrix for $k = 1, \ldots, K$, K the number of clusters, and u_{ik} is the membership degree of the ith sample to the kth cluster. λ is the fuzzy clustering exponent. $d_{ik}^{(wf)}$ is the weighted Euclidean distance between the ith sample and the kth cluster center defined as:

$$d_{ik}^{(wf)} = d^{(wf)}(\mathbf{x}_i, \mathbf{z}_k) = \left[\sum_{m=1}^{M} wf_m (x_{im} - z_{km})^2 \right]^{1/2}, \tag{5}$$

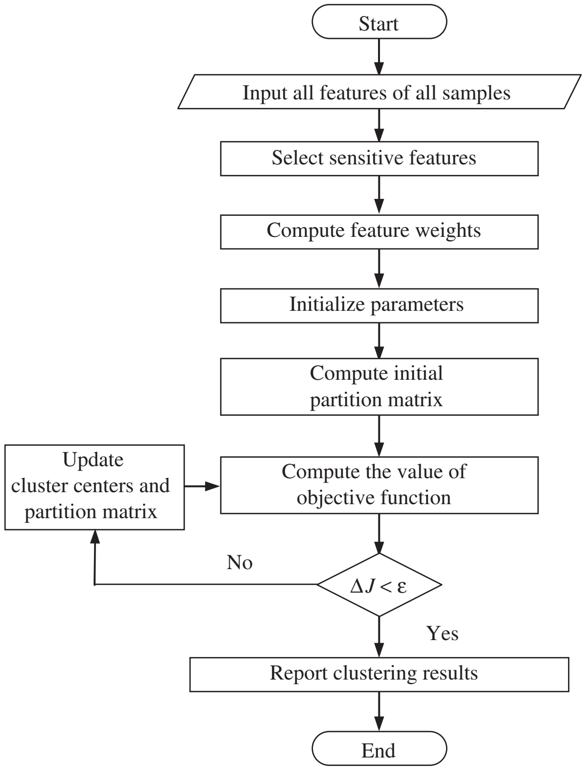

Fig. 4. Flow chart of the new clustering algorithm.

where $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_k, \ldots, \mathbf{z}_K\}$ is the set of cluster centers, $\mathbf{z}_k = (z_{k1}, \ldots, z_{km}, \ldots, z_{kM})$ is the centre of the kth cluster and z_{jm} is its mth component.

Substituting **wf** computed in Section 3.2.2 into Eq. (3), constructing and solving the Lagrange equation, then we derive the computational formulae of $\mathbf{z}_k$ and u_{ik} as:

$$\mathbf{z}_k = \sum_{i=1}^{N} (u_{ik})^{\lambda} \mathbf{x}_i \Big/ \sum_{i=1}^{N} (u_{ik})^{\lambda} \tag{6}$$

and

$$u_{ik} = 1 \Big/ \sum_{a=1}^{K} (d_{ik}^{(wf)} / d_{ia}^{(wf)})^{2/(\lambda-1)}, \tag{7}$$

where $d_{ia}^{(wf)}$ is the weighted Euclidean distance between the ith sample and the ath cluster centre for $a = 1, \ldots, K$.

The procedural steps of the proposed clustering algorithm can be summarised in Table 5.

The choice of the fuzzy clustering exponent λ is similar to the FCM algorithm, i.e. the optimal range of the fuzzy clustering exponent λ is in [1.5, 2.5], and λ is generally set to the middle value 2 [26].

5. Fault diagnosis

5.1. Experiments and results

In this section, three experiments over three different data sets collected with the experimental system of locomotive roller bearings are conducted to demonstrate the effectiveness of the new clustering algorithm.

Table 5
The proposed clustering algorithm

(1) Selecting the sensitive features according to the threshold value of the distance evaluation criterion ϕ, and computing the feature weights using CDET.
(2) Initialising the parameters related to the new clustering algorithm: the number of clusters K, the fuzzy clustering exponent λ, and a threshold value ε.
(3) Given the number of clusters K and the fuzzy clustering exponent λ, initialising fuzzy partition matrix $[u_{ik}]$.
(4) Updating the fuzzy cluster centres $\mathbf{z}_k$ and the fuzzy partition matrix $[u_{ik}]$ using Eqs. (6) and (7).
(5) If the difference ΔJ between two adjacent computed values of the objective function J is larger than the given threshold value ε, then going to Step 4; otherwise going to Step 6.
(6) Reporting the cluster centre set $\mathbf{Z}$ and fuzzy partition matrix $\mathbf{U}$, and ending the algorithm.

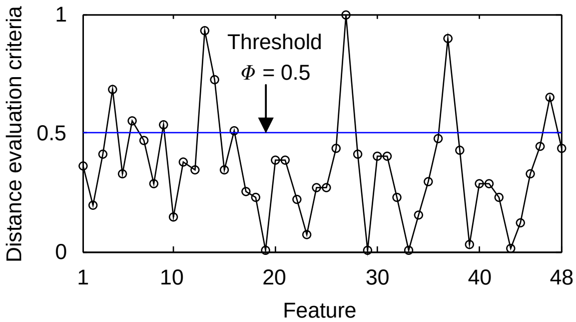

Fig. 5. Distance evaluation criteria of all features.

For comparison, the FCM algorithm (Algorithm 1), and the FCM algorithm with feature selection and no weighting (Algorithm 2) are also employed to analyse the same data sets, respectively. For convenience, the proposed algorithm is referred as Algorithm 3 in the following section.

5.1.1. Clustering performance comparison of different fault categories

In this experiment, a data set consisting of the four data subsets acquired respectively under normal condition, serious fault in the outer race, inner race fault and roller fault of bearings, is applied to evaluate the performance of the proposed algorithm in recognising a single fault of the locomotive roller bearings. As mentioned in Section 2, each of the four data subsets contains 50 samples, and therefore the whole data set corresponding to the four conditions of bearings includes altogether 200 samples.

The distance evaluation criteria $\overline{\alpha}_j$ $(j = 1, 2, \ldots, 48)$ are shown in Fig. 5. The nine sensitive features are selected from the 48 features with CDET when the threshold ϕ is set as 0.5. Fig. 6 shows the feature weights of the nine sensitive features computed using CDET.

For visualisation, principal component analysis (PCA) method is implemented on each of the clustering results produced by the three algorithms. The plots of the first three principal components (PCs) of their results are shown in Figs. 7(a)–(c), respectively. In Figs. 7(a) and (b), many samples (circled by "o") are misclassified by Algorithms 1 and 2, but in Fig. 7(c), using Algorithm 3, almost all samples are correctly classified to the corresponding clusters except that the three samples of inner race fault (circled by "o") are misclassified as normal condition.

5.1.2. Clustering performance comparison of different fault severities

To examine the effect of the new clustering algorithm in identifying incipient faults occurring in the locomotive roller bearings, the proposed algorithm is performed on a data set containing different fault severities. The data set has 150 samples: 50 samples being normal condition, 50 samples for slight fault in the outer race and the remaining 50 belonging to serious fault in the outer race.

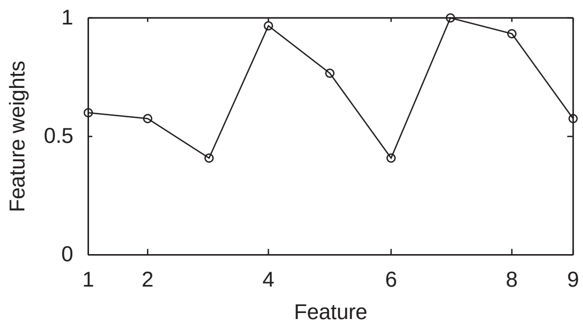

Fig. 6. Feature weights of the sensitive features.

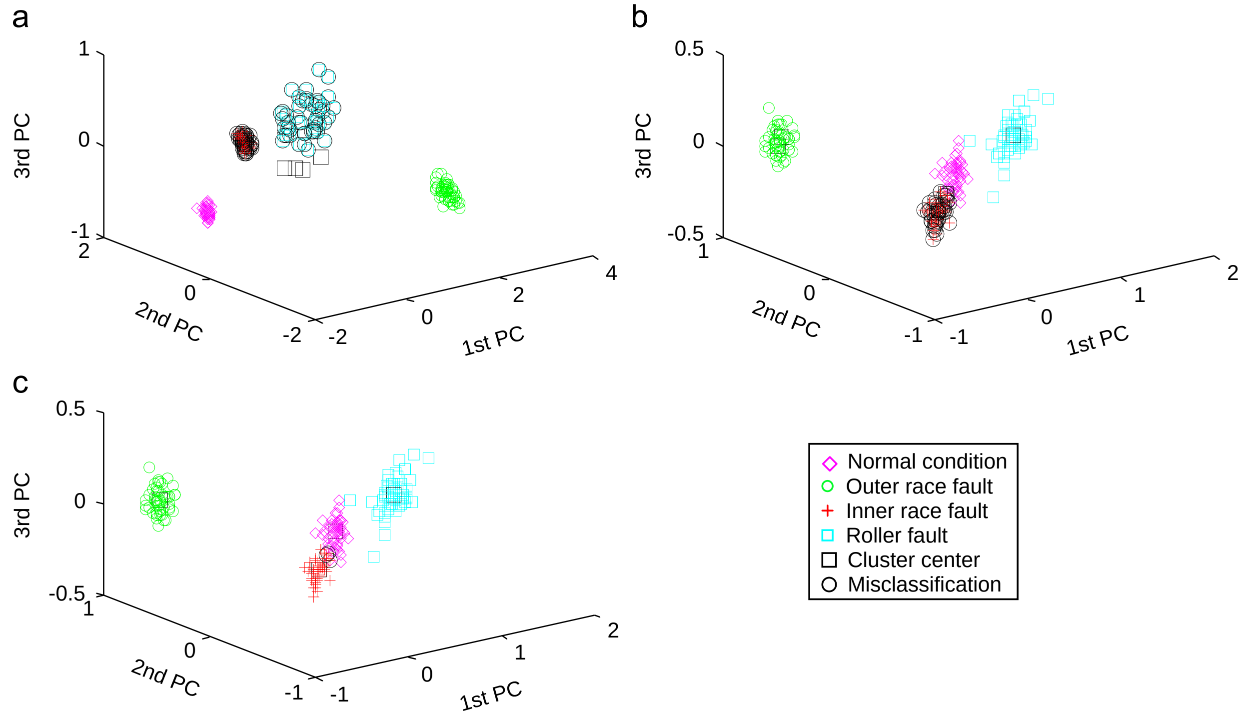

Fig. 7. Scatter plots of principal components for clustering results of experiment 1: (a) Algorithm 1, (b) Algorithm 2, and (c) Algorithm 3.

Fig. 8 displays the distance evaluation criteria of all the features. The feature weights of the selected sensitive features are shown in Fig. 9 when the two-stage feature selection and weighting technique based on CDET is adopted and the threshold ϕ is selected as 0.5.

The plots of the first three PCs of the experimental results obtained by Algorithms 1–3, are shown in Figs. 10(a)–(c), respectively. From the three figures, it can be seen that, a majority of samples of slight fault (circled by "o") are misclassified into normal class with Algorithm 1, and whether using Algorithm 2 or Algorithm 3, all samples are identified to the corresponding clusters accurately. Because of too many samples and the limited space, only five samples are selected randomly from each of the three conditions, and their memberships belonging to each condition using the three clustering algorithms are presented in Table 6, respectively.

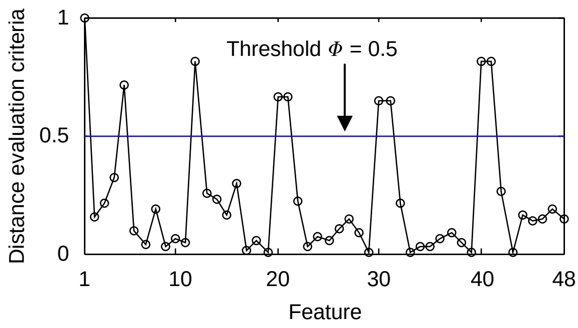

Fig. 8. Distance evaluation criteria of all features.

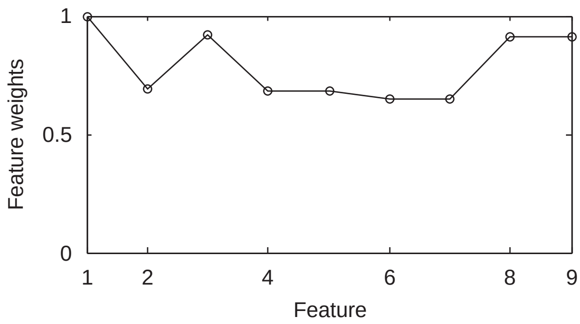

Fig. 9. Feature weights of the sensitive features.

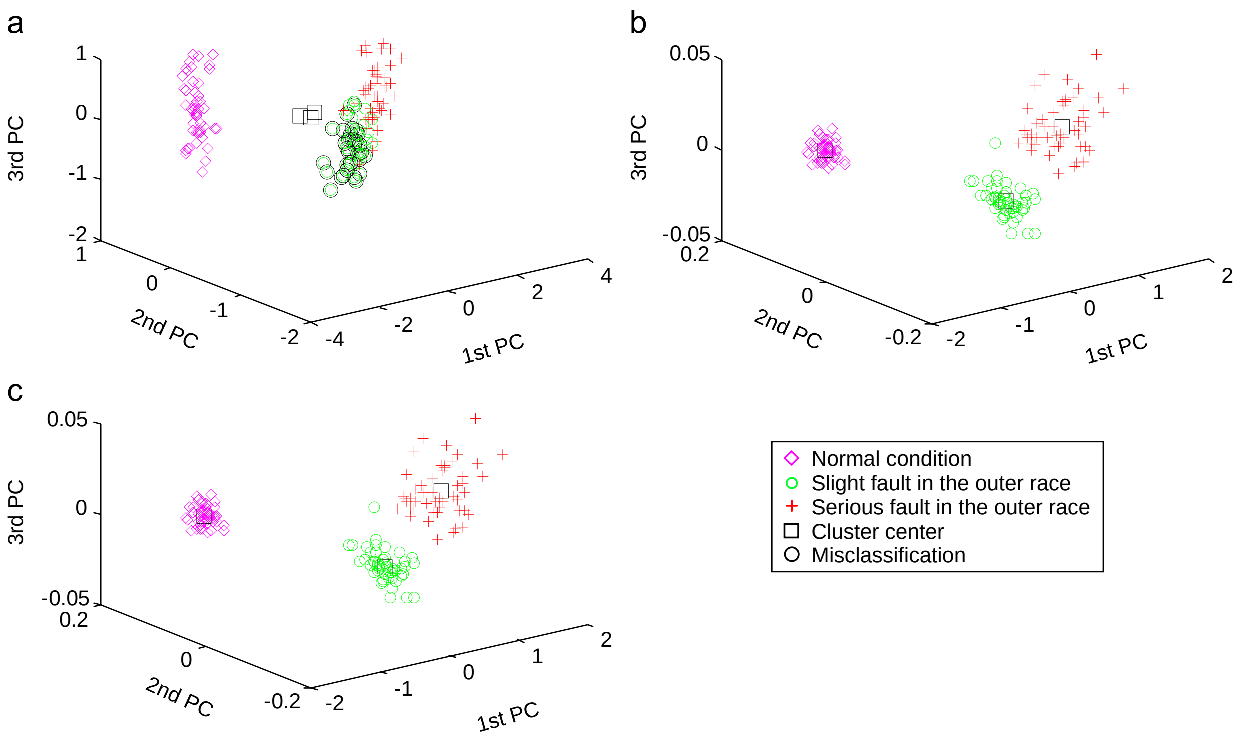

Fig. 10. Scatter plots of principal components for clustering results of experiment 2: (a) Algorithm 1, (b) Algorithm 2, and (c) Algorithm 3.

Table 6
The membership of the 15 samples from the three different conditions

Condition	Sample	Algorithm 1			Algorithm 2			Algorithm 3		
		Normal condition	Slight fault	Serious fault	Normal condition	Slight fault	Serious fault	Normal condition	Slight fault	Serious fault
Normal condition	1	0.38527	0.38278	0.23195	0.99423	0.000378	0.005393	0.99518	0.000402	0.004419
	2	0.37202	0.38252	0.24547	0.99693	0.000264	0.002802	0.99716	0.000306	0.002532
	3	0.36664	0.3732	0.26016	0.99724	0.000188	0.002575	0.99746	0.000218	0.00232
	4	0.38541	0.37652	0.23808	0.98596	0.000833	0.013206	0.98604	0.001043	0.012917
	5	0.39876	0.36872	0.23252	0.99851	0.00012	0.001372	0.99895	0.000105	0.000944
Slight fault	6	0.27429	0.30314	0.42257	0.001705	0.99724	0.001059	0.001412	0.99776	0.000826
	7	0.28259	0.30643	0.41097	0.002716	0.99553	0.001759	0.002312	0.99628	0.001408
	8	0.28548	0.30865	0.40587	0.001283	0.99789	0.000824	0.00124	0.99801	0.00075
	9	0.27583	0.30373	0.42045	0.001285	0.99789	0.000825	0.001141	0.99817	0.000691
	10	0.28054	0.30555	0.41391	0.00165	0.99729	0.001062	0.001488	0.99761	0.000902
Serious fault	11	0.40221	0.34374	0.25405	0.001732	9.560e−5	0.99817	0.001457	9.680e−5	0.99845
	12	0.39453	0.34049	0.26498	0.008438	0.000385	0.99118	0.007865	0.000438	0.9917
	13	0.39248	0.34537	0.26215	0.003963	0.000226	0.99581	0.003549	0.000243	0.99621
	14	0.39485	0.34136	0.26379	0.001895	9.470e−5	0.99801	0.001457	8.880e−5	0.99845
	15	0.39913	0.34174	0.25913	0.008621	0.000397	0.99098	0.007605	0.000428	0.99197

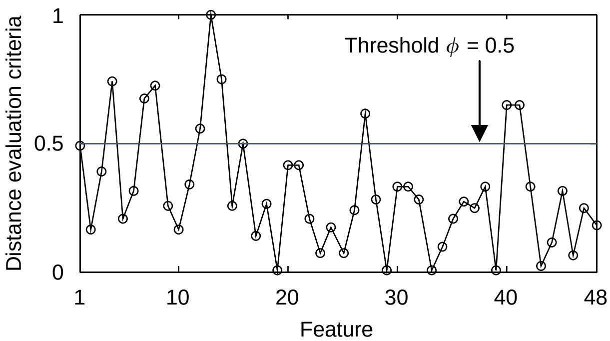

Fig. 11. Distance evaluation criteria of all features.

5.1.3. *Clustering performance comparison of compound faults*

In rotating machinery, besides the different categories of a single fault and the different severities of a fault, generally compound faults also occur. If they cannot be detected as early as possible, they will usually lead to a more serious consequence. Thus, it is of considerable practical value to diagnose the compound faults in rotating machinery correctly and in time.

In this experiment, in order to validate the robustness and practicability of the new algorithm, a data set containing not only the different fault categories and severities but also the compound faults is collected. The data set represents the nine conditions (described in Table 2) of the locomotive roller bearings and consists of 450 samples.

The distance evaluation criteria of the 48 features are presented in Fig. 11. When the threshold ϕ is set as 0.5, the nine sensitive features are selected from the original feature set using feature selection based on CDET. Fig. 12 shows the feature weights of the nine sensitive features obtained via feature weighting based on CDET.

Figs. 13(a)–(c) show the clustering results of the nine conditions of bearings with the three different algorithms, respectively. A lot of samples (circled by "o") are misclassified by Algorithm 1 in Fig. 13(a) and all of the samples of inner race fault are not recognised correctly with Algorithm 2. Most of them are classified as

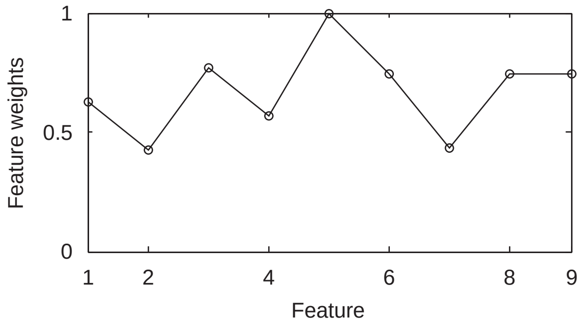

Fig. 12. Feature weights of the sensitive features.

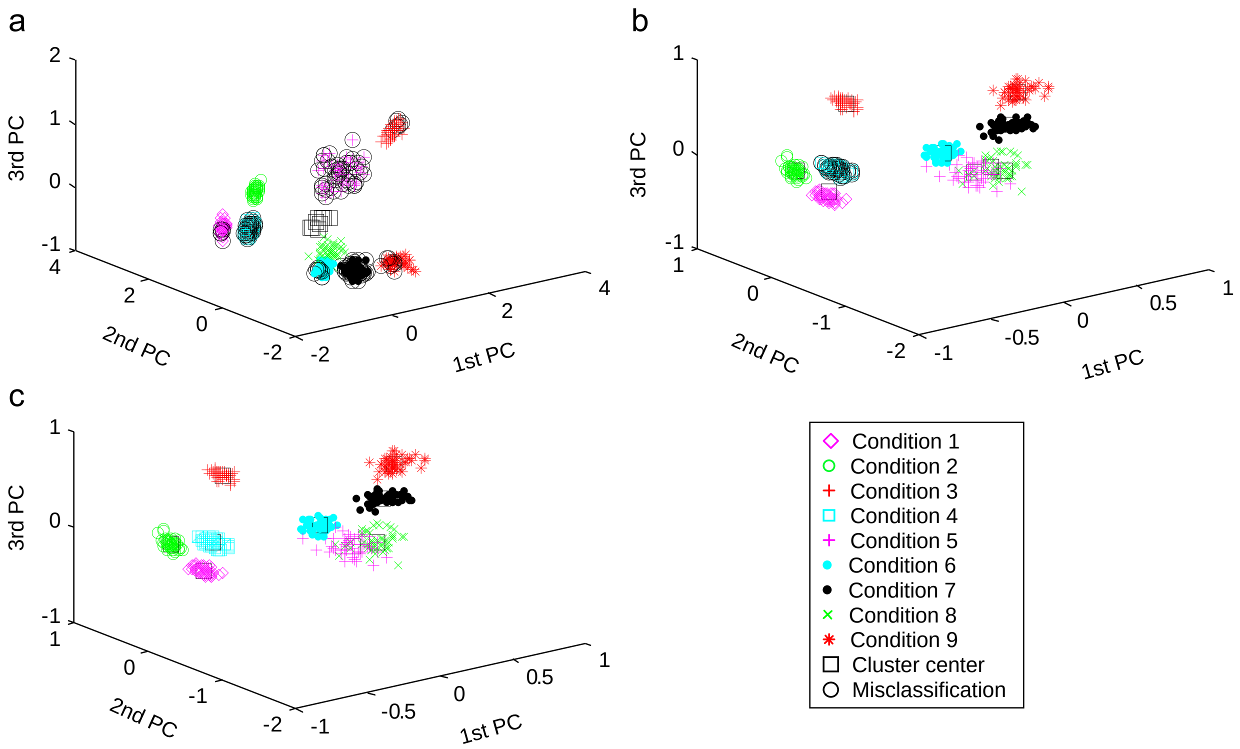

Fig. 13. Scatter plots of principal components for clustering results of experiment 3: (a) Algorithm 1, (b) Algorithm 2, and (c) Algorithm 3.

normal condition and the rest are identified as the condition of slight fault in the outer race. However, utilising Algorithm 3, all samples are correctly diagnosed in Fig. 13(c).

Table 7 gives the number of misclassification, the error rates and the values of the objective function of the three different clustering algorithms in the three experiments. The table shows that the error rates of the three algorithms are, respectively, 50%, 25% and 3% for experiment 1, 22.67%, 0 and 0 for experiment 2, and 41.11%, 11.11% and 0 for experiment 3. From Table 7, it is seen that the value J of the objective function obviously drops from Algorithm 1 to Algorithm 3 in turn for each experiment.

5.2. Discussion

(1) Comparing Algorithm 2 with Algorithm 1, it is found that the clustering accurate rate (75% for experiment 1, 100% for experiment 2, and 88.89% experiment 3) of the former obviously outperforms the

Table 7
The clustering results

	Algorithm 1			Algorithm 2			Algorithm 3		
	Misclassification number	Error (%)	J	Misclassification number	Error (%)	J	Misclassification number	Error (%)	J
Experiment 1	100	50	774.43	50	25	23.14	3	1.5	8.25
Experiment 2	34	22.67	638.95	0	0	1.34	0	0	0.91
Experiment 3	185	41.11	1184.80	50	11.11	63.19	0	0	24.82

latter (50% for experiment 1, 77.33% for experiment 2, and 58.89% experiment 3). For example, in Fig. 7, many samples are misclassified by Algorithm 1. It is because that in fault diagnosis, different features have different importance degrees to identify the different faults. Some features are sensitive and closely related to the faults, but others are not. In Fig. 7(a), some of the features from the original feature set contain too much fault-unrelated information and there is a high degree of overlap between the values of these features between the clusters. These features would confuse the clustering process, and when they are used to diagnose fault, the clustering success will decrease clearly.

(2) Another comparison result between Algorithms 3 and 2 indicates that Algorithm 3 is generally superior to Algorithm 2 in the light of clustering accuracy. The accuracies of Algorithm 3 (97% for experiment 1, 100% for experiment 2 and 100% for experiment 3), respectively, increase by 25%, 0 and 11.11% compared with Algorithm 2 for experiments 1, 2 and 3. These confirm our idea that the proposed clustering algorithm based on CDET may not only select the sensitive features from the original feature set but also indicate their different importance in fault diagnosis. Thus, it may attain the best diagnosis result among the three clustering algorithms.

(3) Because experiment 2 is a simple three-class clustering problem and easy to be solved, the best clustering accuracy is obtained by Algorithms 2 and 3. However, the value J of the objective function of Algorithm 3 is smaller than that of Algorithm 2 in this experiment. Also, from the memberships listed in Table 6, it may be seen that the result with Algorithm 3 is more accurate than the others. The clustering result using Algorithm 1 is extremely vague. Similarly, in the other two experiments, the value J of Algorithm 3 still remains the smallest compared with Algorithms 1 and 2. This implies that Algorithm 3 helps to reduce the vagueness and the uncertainty of clustering and therefore its error rate necessarily decreases.

(4) An interesting observation from the figures of the clustering results of the three experiments is that each of the cluster centres is closer to the dense area of the corresponding cluster after feature selection and weighting. This observation signifies that the proposed two-stage feature selection and weighting may remove the irrelevant features, highlight the importance of the sensitive features and improve the clustering performance.

(5) In all the experiments, the inexact clustering results are yielded by Algorithm 1. By using feature selection based on CDET in Algorithm 2, the improved diagnosis results are attained and even this algorithm may correctly recognise the different fault severities in experiment 2. However, Algorithm 2 suffers on performance when it diagnoses the compound faults, and all of the samples of inner race fault are misclassified with it. But, Algorithm 3 proposed in this paper may reliably recognise not only the different fault categories and severities but also the compound faults. The reason is as follows: the new clustering algorithm believes that the original features have different importance to distinguishing the different conditions of bearings. It is able to automatically acquire the information about the bearing data set, select the sensitive features and assign the feature weights to them to increase the clustering performance.

(6) The problems studied in the three experiments cover single fault diagnosis, incipient fault diagnosis, and compound fault diagnosis, and therefore they are typical diagnosis cases from the view of fault diagnosis. The distribution of the faulty samples includes various shapes, such as hyperspherical, nonspherical, compact and loose clusters. Thus, they are also representative clustering problems. In a word, the satisfied diagnosis results of the three experiments demonstrate the practicability and robustness of the proposed clustering algorithm.

(7) In the present study, the number of clusters is set according to experience. Requiring users to predefine the number of clusters beforehand is a shortcoming of the proposed algorithm as well as the FCM algorithm. We would like to study this problem and overcome it in future.

(8) Through the three experiments of fault diagnosis of locomotive roller bearings, it can be found that two frequency-domain features x_{fc} (frequency centre) and x_{rmsf} (root mean square frequency) extracted not only from FFT spectrums but also the envelope spectrums of their filtered signals, are mostly frequently selected, and x_{mf} (mean frequency), x_{stdf} (standard deviation frequency) and SF(shape factor) are also used frequently. This might suggest that these features are sensitive and closely related to the bearing faults, and they should be calculated in fault diagnosis. The spectrum peak ratios of outer race, inner race, and roller are seldom selected and used in our experiments. It might shows that the spectrum peak ratios of the envelope spectrums of the vibration signals are not always the best features for diagnosing bearing fault, especially for compound fault occurring in the bearings. Although the proposed algorithm is applied to fault diagnosis of the roller bearings successfully, it may also be employed to fault diagnosis of other rotating machinery.

(9) Although CDET and PCA are both used in the paper, they are used for different purposes. Using CDET is to select sensitive features, while using PCA is to present the results and to see the results easily. Of course, both CDET and PCA may be used to reduce features. The selected features by CDET remain physical meaning. However, using PCA to reduce features, we cannot understand the meaning that the PCs represent.

6. Conclusion

In this work, a fault diagnosis method based on a new clustering algorithm is presented, which overcomes the existing shortcoming in the FCM algorithm and takes account of the different importance degrees of features in clustering. First, both time-domain and frequency-domain feature parameters are, respectively, extracted from the raw and filtered vibration signals to recognise the operation conditions of rotating machinery. Then, a two-stage feature selection and weighting technique based on the compensation distance evaluation technique (CDET) are introduced and feature weights are computed via CDET according to sensitivity of features. Finally, it is incorporated into the FCM algorithm to develop the new clustering algorithm. The simple computation of the feature weights makes it much easier to be applied to fault diagnosis.

The new clustering algorithm is applied to incipient fault diagnosis and compound fault diagnosis of locomotive roller bearings. Compared with the other two clustering algorithms, this algorithm shows more powerful ability in identifying the different fault categories, fault severities, and the compound faults. Thus, it is a promising method to fault diagnosis of roller bearings. Of course, the algorithm discussed here can also be applied to fault diagnosis of other rotating machinery.

Acknowledgements

This work is supported by the key project of National Nature Science Foundation of China (No. 50335030), National High-tech R&D Program of China (No. 2006AA04Z430) and National Basic Research Program of China (No. 2005CB724100). Thank Zhuzhou Times Electric Group for its experimental set-up.

References

[1] J. Yang, Y. Zhang, Y. Zhu, Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension, Mechanical Systems and Signal Processing (2006), doi: 10.1016/j. ymssp. 2006.10.005.

[2] J. Sanz, R. Perera, C. Huerta, Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms, Journal of Sound and Vibration (2007), doi: 10.1016/j. jsv. 2007.01.006.

[3] B. Liu, Selection of wavelet packet basis for rotating machinery fault diagnosis, Journal of Sound and Vibration 284 (2005) 567–582.

[4] X.S. Lou, K.A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mechanical Systems and Signal Processing 18 (2004) 1077–1095.

[5] M.L.D. Wong, L.B. Jack, A.K. Nandi, Modified self-organising map for automated novelty detection applied to vibration signal monitoring, Mechanical Systems and Signal Processing 20 (2006) 593–610.

[6] H. Ocak, K.A. Loparo, F.M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling—a method for bearing prognostics, Journal of Sound and Vibration (2007), doi: 10.1016/j. jsv. 2007.01.001.
[7] L.B. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms, Mechanical Systems and Signal Processing 16 (2002) 373–390.
[8] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing 20 (2006) 1483–1510.
[9] W. Yang, Establishment of the mathematical model for diagnosing the engine valve faults by genetic programming, Journal of Sound and Vibration 293 (2006) 213–226.
[10] Z. Long, Neural networks based on fuzzy clustering and its applications in electrical equipment's fault diagnosis, in: Proceedings of IEEE International Conference on Machine Learning and Cybernetics, 2005, pp. 3996–3999.
[11] C. Anpin, L. Changchun, Fuzzy approaches for fault diagnosis of transformers, Fuzzy Sets and Systems 118 (2001) 139–151.
[12] X. Rui, W. Donald, Survey of clustering algorithms, IEEE Transactions on Neural Networks 16 (2005) 645–678.
[13] T.W. Liao, Clustering of time series data—a survey, Pattern Recognition 38 (2005) 1857–1874.
[14] C.D. Joseph, Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems, Journal of Cybernetics 4 (1974) 1–15.
[15] C.B. James, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[16] X.B. Gao, W.X. Xie, Advances in theory and applications of fuzzy clustering, Chinese Science Bulletin 45 (2000) 961–970.
[17] Y.C. Elaine, K.C. Wai, K.N. Michael, Z.H. Joshua, An optimization algorithm for clustering using weighted dissimilarity measures, Pattern Recognition 37 (2004) 943–952.
[18] X. Wang, Y. Wang, L. Wang, Improving fuzzy *c*-means clustering based on feature-weight learning, Pattern Recognition Letters 25 (2004) 1123–1132.
[19] F. Hichem, N. Olfa, Unsupervised learning of prototypes and attribute weights, Pattern Recognition 37 (2004) 567–581.
[20] B.S. Yang, T. Han, J.L. An, ART-KOHONEN neural network for fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing 18 (2004) 645–657.
[21] B.S. Yang, K.J. Kim, Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals, Mechanical Systems and Signal Processing 20 (2006) 403–420.
[22] A. Widodo, B. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mechanical Systems and Signal Processing (2007), doi: 10.1016/j. ymssp. 2006.12.007.
[23] Q. Hu, Z.J. He, Z.S. Zhang, Y.Y. Zi, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble, Mechanical Systems and Signal Processing 21 (2007) 688–705.
[24] B. Samanta, K.R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, Mechanical Systems and Signal Processing 17 (2003) 317–328.
[25] U. Benko, J. Petrovčič, D. Juričič, J. Tavčar, J. Rejec, An approach to fault diagnosis of vacuum cleaner motors based on sound analysis, Mechanical Systems and Signal Processing 19 (2005) 427–445.
[26] R.P. Nikhil, C.B. James, On cluster validity for the fuzzy *c*-means model, IEEE Transactions on Fuzzy Systems 16 (1995) 370–379.