ORIGINAL ARTICLE

Fault diagnosis of rotating machinery based on a new hybrid clustering algorithm

Yaguo Lei · Zhengjia He · Yanyang Zi · Qiao Hu

Abstract A new hybrid clustering algorithm based on a three-layer feed forward neural network (FFNN), a distribution density function, and a cluster validity index, is presented in this paper. In this algorithm, both feature weighting and sample weighting are considered, and an optimal cluster number is automatically determined by the cluster validity index. Feature weights are learnt via FFNN based on the gradient descent technique, and sample weights are computed by using the distribution density function of a sample. Feature weighting and sample weighting highlight the importance of sensitive features and representative samples, and simultaneously weaken the interference of insensitive features and vague samples. The presented algorithm is described and applied to the incipient fault diagnosis of locomotive roller bearings. The diagnosis result demonstrates the superior effectiveness and practicability of the algorithm, and shows that it is a promising approach to the fault diagnosis of rotating machinery.

Keywords Feature weighting · Sample weighting · Cluster validity index · Hybrid clustering · Fault diagnosis

Y. Lei (✉) · Q. Hu
Department of Mechanical Engineering,
Xi'an Jiaotong University,
Xi'an 710049, People's Republic of China
e-mail: leiyaguo@163.com

Z. He · Y. Zi
State Key Laboratory for Manufacturing Systems Engineering,
Xi'an Jiaotong University,
Xi'an 710049, People's Republic of China

1 Introduction

Rotating machinery is critical equipment in key industries, such as power plants, petrochemical plants, etc. Faults in rotating machinery can cause both personal casualties and economical loss [1]. Therefore, the fault diagnosis of rotating machinery has become a vigorous area of study and has attracted more and more attention during the past decade. The fault diagnosis of rotating machinery consists of three key steps: data acquisition, feature extraction, and fault detection and identification [2–4]. It is a procedure of mapping the information obtained in the measurement space and/or features in the feature space to machine faults in the fault space. Therefore, it is essentially a problem of pattern recognition and pattern classification [5, 6]. With the development of artificial intelligence, such as artificial neural networks (ANNs), fuzzy sets theory, and expert systems, etc., many fault diagnosis approaches have emerged as new techniques for fault diagnosis systems [7–9].

Fuzzy clustering, owing to its superiority in dealing with uncertainty and independence from supervisors, has been widely studied and applied to fault diagnosis [10–12]. 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 [13, 14].

Among various clustering algorithms, the fuzzy c-means (FCM) algorithm [15, 16] is one of the most well-known and widely used algorithms [17]. However, the FCM

algorithm suffers from the following shortcomings: (1) it supposes that all features have uniform contributions to clustering, and does not take account of their different importance degrees; (2) it always treats each sample equally, and neglects their different influences to clustering; (3) it requires users to predefine the number of clusters beforehand, but it is not always possible to know the number in advance. 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. Chan et al. [18] 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. [19] introduced a weighted FCM algorithm based on weighted Euclidean distance. Frigui and Nasraoui [20] presented a new approach to perform clustering and feature weighting simultaneously, and used it to segment color images. To solve the problem of predefining the number of clusters in the FCM algorithm beforehand, Sun et al. [21] developed a new FCM-based clustering algorithm and a new index for validating clustering results. In addition, many cluster validity indexes have been developed, by which an optimal cluster number can be automatically found [22–27].

In all of the clustering algorithms mentioned above, no researcher simultaneously solved the above three problems of the FCM algorithm. Thus, we make much improvement and propose a new hybrid clustering algorithm to overcome the three shortcomings. The proposed algorithm is developed by incorporating feature weighting, sample weighting, and a cluster validity index into the FCM algorithm, and is, finally, applied to the incipient fault diagnosis of rotating machinery. The optimal cluster number is automatically determined by using a cluster validity index. Feature weights are learnt via a three-layer feed forward neural network (FFNN) under the unsupervised training mode, and are assigned to the corresponding features to indicate their different sensitivities. Sample weights are computed by using a distribution density function of a sample and are then assigned to the relevant samples to highlight the effects of the representative samples.

In comparison with the existing algorithms [18–21], the merits of the proposed algorithm are not only to automatically find the optimal cluster number, but also to simultaneously take account of feature weighting and sample weighting. Thus, it has better clustering performance than the others.

This paper is organized as follows. Section 2 describes the computation of the feature weights and the sample weights, and presents the definition of the cluster validity index. Section 3 presents the hybrid clustering algorithm. The proposed algorithm is developed by combining the

FCM algorithm with feature weighting, sample weighting, and the cluster validity index. In Section 4, the proposed algorithm is applied to the incipient fault diagnosis of the locomotive roller bearings and compared with the FCM algorithm in clustering performance. Our conclusions are drawn in Section 5.

2 Computing weights and choosing a cluster validity index

2.1 Motivation

Figure 1 shows a simple example that illustrates the necessity of feature weighting and sample weighting. Figure 1a describes a data set containing three clusters in a three-dimensional space represented by features f1, f2, and f3, respectively. Each cluster consists of 12 samples and they are situated on two concentric spherical surfaces, respectively. The center of the concentric spherical surfaces is that of the corresponding cluster. Figure 1b–d show the projections of the data set on different planes, respectively.

From Fig. 1a, it is easily seen that not all samples in each cluster have the same importance to clustering. The six samples situated on the inner spherical surface are more important to clustering than those situated on the outer one, i.e., the former are more representative to clustering. By
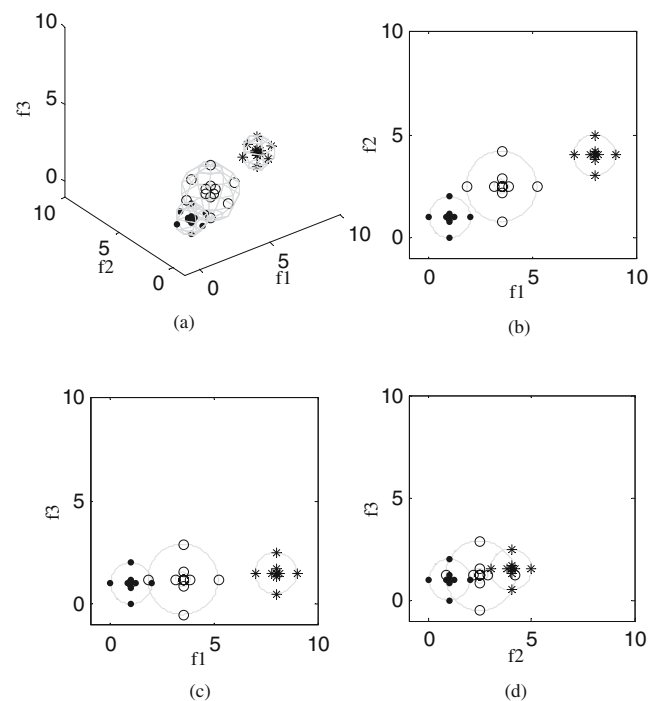

Fig. 1a–d A data set with three clusters and its projections on different planes. **a** A data set distributed in a three-dimensional space. **b** Projection of the data set on the f1–f2 plane. **c** Projection of the data set on the f1–f3 plane. **d** Projection of the data set on the f2–f3 plane

comparing Fig. 1b-d, one can see that each feature has different importances to clustering. Obviously, f1 is superior to the others to discriminate these three clusters, f2 is inferior, and f3 is the worst. Thus, to improve clustering performance, both feature weighting and sample weighting are absolutely necessary.

2.2 Learning feature weights based on a three-layer FFNN

Feature weighting can be regarded as a generalization of feature selection, and it is a number within [0, 1] that 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 to shrink the axes corresponding to the features unrelated to clustering. Consequently, it enables a clustering algorithm to classify nonspherical clusters as well as hyperspherical ones.

The feature weights in this paper are learnt via a three-layer FFNN, as shown in Fig. 2. They are learnt based on the gradient descent technique through minimizing the following objective function with respect to the feature weight vector $\mathbf{wf}$ [28, 29]:

$$E(\mathbf{wf}) = 2/[N(N-1)] \sum_{j<i} \left[\rho_{ij}^{(wf)}\left(1 - \rho_{ij}^{(1)}\right) + \rho_{ij}^{(1)}\left(1 - \rho_{ij}^{(wf)}\right) \right]\bigg/2 \tag{1}$$

where $\mathbf{wf} = (wf_1,..., wf_m,..., wf_M)$ denotes the feature weight vector. wf_m is the mth feature weight for $m=1,..., M$ and M is the number of features of each sample. N is the number of samples of a data set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_N\}$. $\mathbf{x}_j = (x_{j1},..., x_{jm},..., x_{jM})$ is the jth sample for $j=1,..., N$, and x_{jm} is its mth feature. $\rho_{ij}^{(1)}$ represents the similarity measure using the standard Euclidean distance $d_{ij}^{(1)}$. $\rho_{ij}^{(wf)}$ denotes the similarity measure using the weighted Euclidean distance $d_{ij}^{(wf)}$.

The weighted Euclidean distance between the ith sample $\mathbf{x}_i$ and the jth sample $\mathbf{x}_j$ is defined as:

$$d_{ij}^{(wf)} = d^{(wf)}(\mathbf{x_i}, \mathbf{x_j}) = \left[\sum_{m=1}^{M} wf_m \left(x_{im} - x_{jm}\right)^2 \right]^{1/2} \tag{2}$$

where wf_m indicates the sensitivity of the mth feature. When all feature weights are equal to 1, the weighted Euclidean distance $d_{ij}^{(wf)}$ defined by Eq. 2 becomes the standard Euclidean distance $d_{ij}^{(1)}$.

Then, $\rho_{ij}^{(wf)}$ is defined as:

$$\rho_{ij}^{(wf)} = 1\bigg/\left(1 + \beta \cdot d_{ij}^{(wf)}\right), \tag{3}$$

where β is a positive parameter determined by solving the following equation:

$$\frac{2}{N(N-1)} \sum_{j<i} \rho_{ij}^{(1)} = \frac{2}{N(N-1)} \sum_{j<i} \frac{1}{1 + \beta \cdot d_{ij}^{(1)}} = 0.5. \tag{4}$$

The motivation of minimizing the objective function defined by Eq. 1 to learn the feature weights can be described as follows. The objective function will attain its minimum when all similarity measures are close to either 0 or 1. A good partition should have the following property: the samples within one cluster are close to the center and different centers are more separate, which implies that the samples within one cluster are more similar $\left(\rho_{ij}^{(wf)} \to 1\right)$ and dissimilar samples are more separate $\left(\rho_{ij}^{(wf)} \to 0\right)$, so that the value of the objective function is low. Therefore, we hope that, by adjusting the feature weights wf, the original similar objects $\left(\rho_{ij}^{(1)} > 0.5\right)$ become more similar $\left(\rho_{ij}^{(wf)} \to 1\right)$, and the original dissimilar objects $\left(\rho_{ij}^{(1)} < 0.5\right)$ are more separate $\left(\rho_{ij}^{(wf)} \to 0\right)$ [29].

Figure 2 shows the FFNN architecture consisting of three layers with $2M$:M:2 nodes for an input, a hidden, and an output layer, respectively. The $2M$ nodes in the input layer correspond to the $2M$ features of any two samples from the data set. The input layer just transfers all inputs to the hidden layer and performs no calculation. The input of the mth node in the hidden layer is the sum of the weighted outputs of the mth and the $(m+M)$th input nodes via connection weights +1 and −1, respectively. The input of the node computing $\rho_{ij}^{(wf)}$ in the output layer is the sum of the weighted outputs of each hidden node with the corresponding weight wf_m, whereas that computing $\rho_{ij}^{(1)}$ is the sum of the weighted outputs of all of the hidden nodes with weights +1. During training, any two samples as a pair are fed to the FFNN. A data set consisting of N samples has $N(N-1)/2$ different sample pairs [28].

Let $in_m^{(l)}$ and $out_m^{(l)}$ denote the input and the output of the mth node($m=1,..., M$) in the lth layer ($l=1, 2, 3$), respectively. We assume that the pairs fed to the FFNN
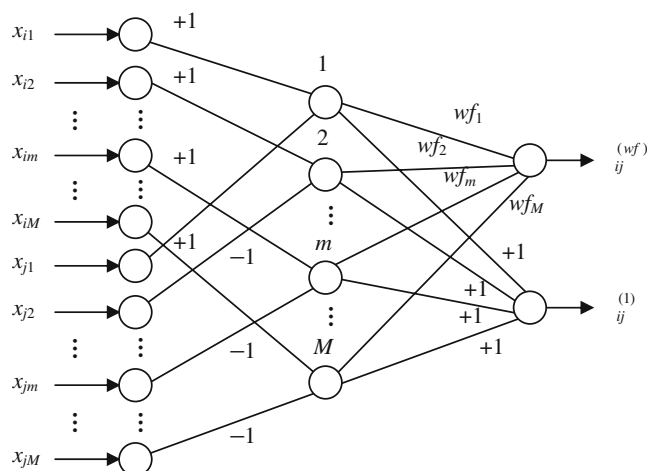

Fig. 2 A three-layer feed forward neural network (FFNN) used to learn feature weights

are $(\mathbf{x}_i, \mathbf{x}_j)$; then, the inputs, the activation functions, and the outputs in each layer are as shown in Table 1.

Substituting the outputs $\rho_{ij}^{(wf)}$ and $\rho_{ij}^{(1)}$ of the FFNN into Eq. 1, we obtain the value of the objective function $E(\mathbf{wf})$. The feature weights are learnt based on the gradient descent technique through minimizing the objective function, where the weight change Δwf_m can be expressed as:

$$\Delta wf_m = -\eta \frac{\partial E(\mathbf{wf})}{\partial wf_m} = -\eta \frac{\partial E}{\partial out_1^{(3)}} \cdot \frac{\partial out_1^{(3)}}{\partial in_1^{(3)}} \cdot \frac{\partial in_1^{(3)}}{\partial wf_m}, \quad (5)$$

$$\frac{\partial E}{\partial out_1^{(3)}} = \frac{2}{N(N-1)} \cdot \sum_{j<i} \frac{1}{2}\left(1 - 2out_2^{(3)}\right), \quad (6)$$

$$\frac{\partial out_1^{(3)}}{\partial in_1^{(3)}} = \frac{\beta}{2\left(1 + \beta \cdot \left(in_1^{(3)}\right)^{1/2}\right)^2 \cdot \left(in_1^{(3)}\right)^{1/2}} \quad (7)$$

$$\frac{\partial in_1^{(3)}}{\partial wf_m} = out_m^{(2)}, \quad (8)$$

where η is the learning rate whose optimal value for each epoch is acquired by using the golden section search method. In each training epoch of the FFNN, the golden section search method is utilized to find the learning rate that minimizes the objection function $E(\mathbf{wf})$ in the given range of the learning rate. The learning rate corresponding to the minimum of the objection function is chosen as the learning rate of the current training epoch.

For all sample pairs of a data set, the FFNN performs forward calculations to obtain the value of the objective function. Then, the feature weight wf_m is iteratively updated with $wf_m + \Delta wf_m$ for each m until the terminating conditions are satisfied, where Δwf_m is computed in accordance with Eqs. 5–8.

Therefore, this FFNN is different from supervised learning neural networks based on input–output pairs and it is trained under the unsupervised mode. The categories of

the training samples are unknown in advance. Therefore, the ANN outputs are not the categories of the input samples, but, instead, are the similarity measures $\rho_{ij}^{(1)}$ and $\rho_{ij}^{(wf)}$ between the input samples. The training purpose is to minimize the objective function $E(\mathbf{wf})$ by adjusting the feature weights. The training process includes the forward pass and the backward pass. During the forward pass, the node outputs advance until the output layer. The backward pass uses the gradient descent method to update the feature weights wf_m.

2.3 Computing sample weights based on a distribution density function

Generally, the more surrounding of other samples that one sample is, then the larger the distribution density of that sample. Then, it is closer to a cluster center and more representative. Therefore, we employ a distribution density function of a sample to compute the sample weights.

The distribution density function y_i of the ith sample ($i=1,..., N$) is defined as:

$$y_i = \sum_{j=1, j\neq i}^{N} \left(1 \big/ d_{ij}^{(wf)}\right) \quad (9)$$

subject to:

$$d_{ij}^{(wf)} \leq r, \quad (10)$$

where r is the boundary of the distribution density function. This boundary is equal to the weighted Euclidean distance where the probability distribution function of $d_{ij}^{(wf)}$ attains its maximum value.

Normalizing y_i according to Eq. 11, we derive the sample weight vector $\mathbf{wp}=(wp_1,..., wp_i,..., wp_N)$. Its component wp_i is the ith sample weight for $i=1,..., N$ and indicates the typicality of the ith sample:

$$wp_i = y_i \bigg/ \sum_i^N y_i. \quad (11)$$

Table 1 Inputs, activation functions, and outputs of the feed forward neural network (FFNN)

Layer	Input	Activation function	Output
$l=1$	$in_m^{(1)} = x_{im}, \; in_{m+M}^{(1)} = x_{jm}$	$out=in$	$out_m^{(1)} = in_m^{(1)}, \; out_{m+M}^{(1)} = in_{m+M}^{(1)}$
$l=2$	$in_m^{(2)} = (+1) \cdot out_m^{(1)} + (-1) \cdot out_{m+M}^{(1)}$	$out=in^2$	$out_m^{(2)} = \left(in_m^{(2)}\right)^2$
$l=3$	$in_1^{(3)} = \sum_m \left(wf_m \cdot out_m^{(2)}\right),$	$out = \frac{1}{(1+\beta \cdot in^{1/2})}$	$out_1^{(3)} = \rho_{ij}^{(wf)} = \frac{1}{\left[1 + \beta \cdot \left(in_1^{(3)}\right)^{1/2}\right]},$
	$in_2^{(3)} = \sum_m out_m^{(2)}$		$out_2^{(3)} = \rho_{ij}^{(1)} = \frac{1}{\left[1 + \beta \cdot \left(in_2^{(3)}\right)^{1/2}\right]}$

2.4 Choosing a cluster validity index

Because the clustering results generally depend on the choice of the cluster number, it is necessary to validate each of the clusters once they are found. This validation is carried out by a cluster validity index, which is to determine an optimal cluster number that can validate the correct and natural description of the data structure. Some numerical examples including several artificial data sets and some UCI databases have been used to compare *PC* [30], *PE* [31], *FS* [32], *MPC* [33], *XB* [34], *FHV* [35], *K* [36], *PCAES* [25], *OS* [26], and *PBMF* [27] indexes. The comparison results show that the *PBMF* index is superior to the others and, therefore, it is chosen as the cluster validity index in this paper. It is defined as [27]:

$$PBMF(K) = \left[\frac{1}{K} \cdot \frac{E_1}{J_\lambda} \cdot D_K \right]^2 \qquad (12)$$

where K is the number of clusters, J_λ is a compactness measure of K clusters, D_K is a separation measure of K clusters, and E_1 is the value of J_λ when $K=1$.

The definition of *PBMF* indicates that larger values of *PBMF* correspond to good clusters, and the number of clusters that maximizes *PBMF* is taken as the optimal number of clusters. More information about *PBMF* can be found in [27].

3 The proposed hybrid clustering algorithm

The proposed hybrid clustering algorithm is created by combining the FCM algorithm with feature weighting, sample weighting, and the cluster validity index mentioned in the preceding section. It consists of an inner iterative loop and an outer iterative loop. The inner iterative loop, similar to the FCM algorithm, performs clustering by minimizing the following objective function. The outer iterative loop computes the validity index of clusters produced by the inner iterative loop. The flow chart of the proposed algorithm is shown in Fig. 3.

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

$$J(\mathbf{U}, \mathbf{Z}, \mathbf{wp}, \mathbf{wf}; \mathbf{X}) = \sum_{k=1}^{K} \sum_{i=1}^{N} wp_i(u_{ik})^\lambda \left(d_{ik}^{(wf)} \right)^2, \qquad (13)$$

subject to:

$$\sum_{k=1}^{K} u_{ik} = 1 \quad 0 \le u_{ik} \le 1, \qquad (14)$$

where $\mathbf{U}=[u_{ik}]$ is the fuzzy partition matrix for $k=1,...,K$ and $i=1,...,N$, and u_{ik} is the membership degree of the ith sample to the kth cluster. λ is the fuzzy clustering exponent.
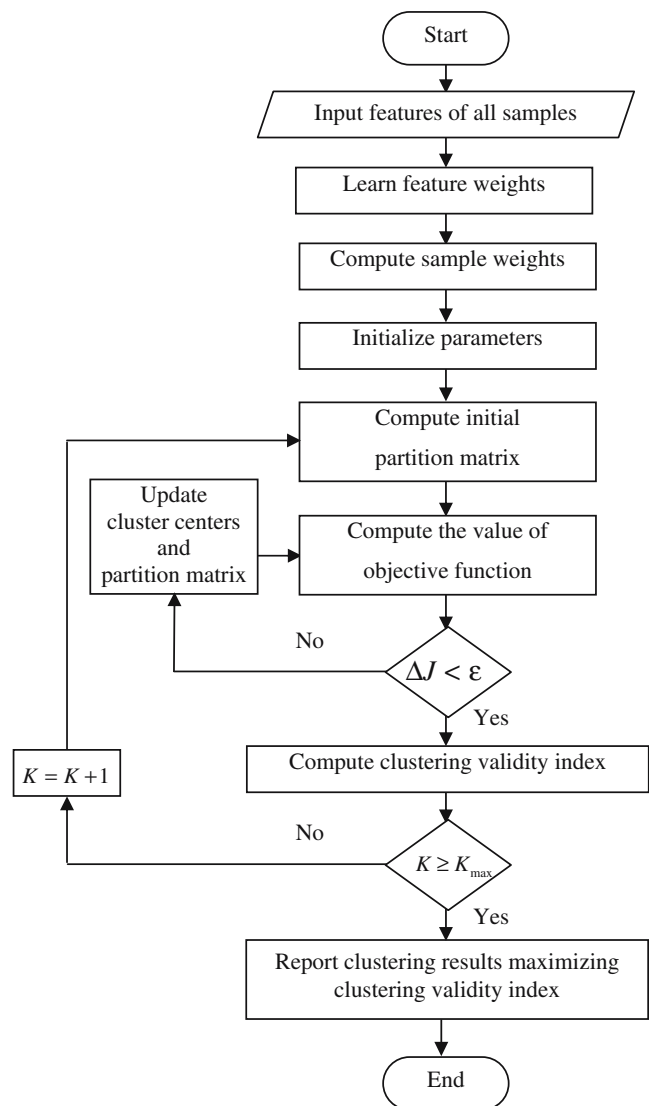

Fig. 3 Flow chart of the proposed hybrid clustering algorithm

$d_{ik}^{(wf)}$ is the weighted Euclidean distance between the ith sample and the kth cluster center, and is defined as:

$$d_{ik}^{(wf)} = d^{(wf)}(\mathbf{x_i}, \mathbf{z_k}) = \left[\sum_{m=1}^{M} wf_m (x_{im} - z_{km})^2 \right]^{1/2}, \qquad (15)$$

where $Z = \{Z_1, \cdots, z_k, \cdots, \cdots z_K\}$ is the set of cluster centers, $\mathbf{z_k}=(z_{kl},..., z_{km},..., z_{kM})$ is the center of the kth cluster and z_{jm} is its mth component.

Substituting $\mathbf{wf}$ introduced in Section 2.2 and $\mathbf{wp}$ computed in Section 2.3 into Eq. 13, constructing and solving the Lagrange equation, then we derive the computational formulae of $\mathbf{z}_k$ and u_{ik} as:

$$\mathbf{z}_k = \sum_{i=1}^{N} wp_i(u_{ik})^\lambda \mathbf{x_i} \Big/ \sum_{i=1}^{N} wp_i(u_{ik})^\lambda \qquad (16)$$

and

$$u_{ik} = 1 \Big/ \sum_{a=1}^{K} \left(d_{ik}^{(wf)} \Big/ d_{ia}^{(wf)} \right)^{2/(\lambda-1)} \qquad (17)$$

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

The procedural steps are as follows:

1. Learn the feature weights via the three-layer FFNN with the gradient descent technique, and compute the sample weights based on the distribution density function, respectively.
2. Initialize the parameters related to the hybrid clustering algorithm: the initial number of clusters, the maximum number of clusters K_{max}, the fuzzy clustering exponent λ, and a threshold value ε.
3. Given the number of clusters K and the fuzzy clustering exponent λ, initialize fuzzy partition matrix $[u_{ik}]$.
4. Update the fuzzy cluster centers z_k and the fuzzy partition matrix $[u_{ik}]$ using Eqs. 16 and 17.
5. If the difference ΔJ between two adjacent computed values of the objective function J is larger than the given threshold ε, then go to Step 4; otherwise, go to Step 6.
6. Compute the *PBMF* index of clusters derived in Step 4.
7. If $K<K_{max}$, then $K=K+1$ and go to Step 3; otherwise, go to Step 8.
8. Find the maximum of *PBMF* and report the value of K that maximizes *PBMF* as the optimal number of clusters. The corresponding cluster center set $\mathbf{Z}$ and fuzzy partition matrix $\mathbf{U}$ are just the clustering results.

The initial number of clusters is 2. According to a rule of thumb that many investigators adopt, the value of K_{max} should not exceed $\sqrt{N}$(N is the number of samples) [37]. 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 of 2 [38].

4 Application to fault diagnosis

4.1 Case study description

Roller bearings, as important components, are widely used in rotating machinery. Roller bearing faults are one of the foremost causes of failures in rotating machinery [33]. Among all types of the locomotive roller bearing faults, including outer race fault, inner race fault, roller fault, and cage fault, etc., the outer race fault occurs most frequently. The incipient faults occurring in the outer races of the locomotive roller bearings generally behave as slight rubbing. If they are not detected as early as possible, they will usually evolve into serious flaking faults, which may lead to fatal breakdowns of the locomotives.

4.2 Data collection

In the fault diagnosis systems of rotating machinery, the most successful method is based on vibration analysis [4, 39, 40]. Here, accelerometers with a bandwidth up to 5,000 Hz are mounted on the bearing housings to acquire the vibration signals from the bearings. The Advanced Data Acquisition and Analysis System by Sony EX are used to collect the data with a sampling frequency of 12.8 KHz.

A bearing data set containing three subsets is obtained from the locomotive roller bearings under three different conditions: (a) normal conditions (NC); (b) with slight rub faults occurring in the outer races (SRF); and (c) with serious flaking faults occurring in the outer races (SFF). Each data subset corresponds to one of the three conditions and it consists of 50 samples. Each sample contains 8,192 sampling points, i.e., the data length of a sample is 8,192.

4.3 Selection and extraction of features

The diagnosis task of machinery is actually a problem of pattern classification and pattern recognition [5, 6], of which, the crucial step is feature selection and extraction. Large kinds of feature parameters have been defined in the pattern recognition field. Here, six of them, which are usually used for the fault diagnosis of roller bearings in the time domain, are shown as follows:

Root mean square (RMS):

$$RMS = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (s_t)^2}, \qquad (18)$$

where s_t ($t=1,..., T$) is the tth sampling point of sample S. T is the number of sampling points.

Peak value (*PV*):

$$PV = \max|s_t|. \qquad (19)$$

Kurtosis value (*KV*):

$$KV = \frac{\sum_{t=1}^{T} (s_t - \bar{s})^4}{(T-1)\sigma^4}, \qquad (20)$$

where $\bar{s}$ is the mean value of sample S, defined as:

$$\bar{s} = \frac{1}{T} \sum_{t=1}^{T} s_t, \qquad (21)$$

and σ is the standard deviation of sample S, defined as:

$$\sigma = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(s_t - \overline{s})^2}. \qquad (22)$$

Crest factor (CF):

$$CF = \frac{\max|s_t|}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(s_t)^2}}. \qquad (23)$$

Impulse factor (IF):

$$IF = \frac{\max|s_t|}{\frac{1}{T}\sum_{t=1}^{T}|s_t|}. \qquad (24)$$

Clearance factor (CLF):

$$CLF = \frac{\max|s_t|}{\left(\frac{1}{T}\sum_{t=1}^{T}\sqrt{|s_t|}\right)^2}. \qquad (25)$$

These six features derive from the amplitude the probability density function of the bearing vibration signals, and they can be calculated and compared with those of normal bearings [2]. It was shown [41] that the peak value and root mean square reflect the vibration amplitude and energy, and the kurtosis value, crest factor, impulse factor, and clearance factor characterize the impact existing in the roller bearings. The kurtosis value and crest factor are robust to varying operating conditions of the bearings, and are good indicators of incipient faults. The impulse and clearance factors are also good indicators of spikiness of the sharp impulses generated by the contact of a defect with the bearing mating surfaces [2, 42]. Faults that typically occur in locomotive roller bearings are usually caused by local defects in the outer race. Such defects generate a series of impact vibrations and cause the vibration amplitude to increase. Therefore, we choose these six features to represent the bearing conditions. But the importance of the six features is different in the fault diagnosis of the bearings. Thus, we assign feature weights to the corresponding features to indicate their different sensitivities in the fault diagnosis of the bearings.

4.4 Diagnosis result

Figure 4 presents the variation of the *PBMF* index with the number of clusters in the range [2, 12] when the proposed algorithm is performed for clustering on the bearing data set. In Fig. 4, it is evident that the maximum exists at $K=3$,
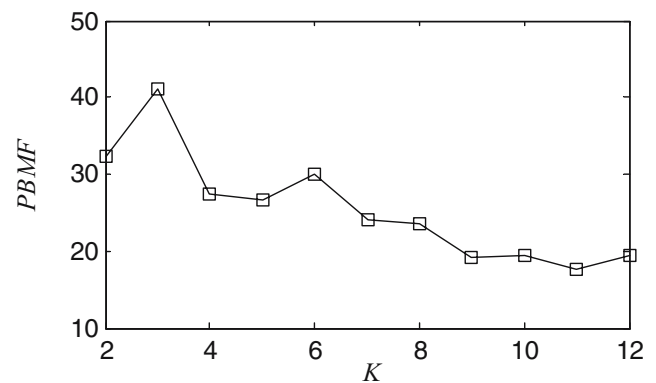

Fig. 4 Variation of the *PBMF* index with the number of clusters

which indicates that the bearing data set consists of three clusters. It is coincident with the fact.

For visualization, we implement the principal component analysis (PCA) method on the clustering results obtained by the proposed algorithm for the bearing data set. The plot of the first two principal components (PCs) of the results is shown in Fig. 5. The two PCs account for more than 97% of the variability of the features. Similarly, Fig. 6 shows the clustering results of the FCM algorithm with the cluster number predefined to 3. In Fig. 6, some samples (circled by an "o") are misclassified by the FCM algorithm, but in Fig. 5, using the proposed algorithm, all samples are correctly classified to the corresponding clusters.

Table 2 gives the values of the objective function, the error rates, and the CPU processing times of the proposed algorithm and the FCM algorithm. Table 2 shows that the error rates of the FCM algorithm and the proposed algorithm are 10% and 0%, respectively. The values of the objective function of the two methods are 16 and 1.2, respectively. The CPU times taken to carry out the FCM
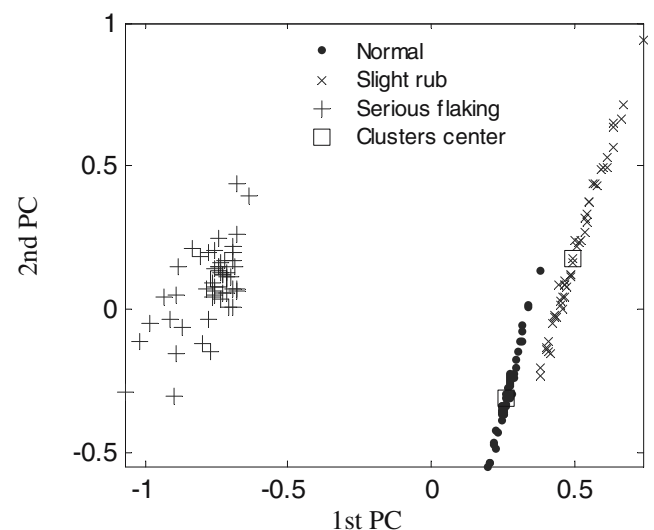

Fig. 5 Scatter plot of the principal components (PCs) for clustering results obtained by the hybrid clustering algorithm
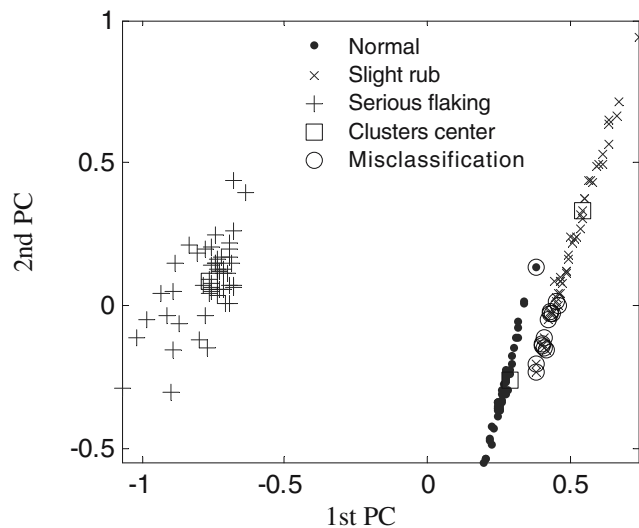
Fig. 6 Scatter plot of the PCs for clustering results obtained by the fuzzy *c*-means (FCM) algorithm

and the proposed algorithms are 0.047 and 2.765 s, respectively.

Comparing the proposed algorithm with the FCM algorithm, we found that the former automatically determines the number of clusters, and, in addition, its clustering accurate rate (100%) obviously outperforms the latter (90%). Although the FCM algorithm diagnoses the serious flaking faults well, it suffers on performance when it diagnoses the incipient slight rub faults. Fifteen samples of the incipient slight rub faults and the normal conditions are misclassified by the FCM algorithm. However, the hybrid clustering algorithm proposed in this paper can diagnose not only the serious flaking faults, but also the incipient slight rub faults perfectly. The reasons for this are as follows. Firstly, the proposed algorithm introduces the cluster validity index into the clustering algorithm. Through utilizing *PBMF* in the clustering algorithm, the best clustering number can be found from the possible values of the number of clusters objectively and automatically.

Table 2 Comparison between the hybrid clustering algorithm and the FCM algorithm

	FCM algorithm			Hybrid clustering algorithm		
	NC	SRF	SFF	NC	SRF	SFF
NC	49	14	0	50	0	0
SRF	1	36	0	0	50	0
SFF	0	0	50	0	0	50
Error (%)	10			0%		
J	16.0			1.2		
CPU time (s)	0.047			2.765		

Secondly, the hybrid clustering algorithm believes that the six time-domain features and the 50 samples of each cluster have different importances to distinguishing these three conditions of the bearings. It is able to automatically acquire the information about the bearing data set and assign the feature weights to the different features and sample weights to the different samples to improve the clustering performance. From Table 2, it is seen that the value *J* of the objective function obviously drops by applying the proposed algorithm to the bearing data set. This implies that the proposed algorithm helps to reduce the vagueness and the uncertainty of clustering and, therefore, its error rate necessarily decreases.

An interesting observation from Figs. 5 and 6 is that each of the three cluster centers in Fig. 5 is closer to the dense area of the samples than that in Fig. 6. This observation signifies that the cluster centers of the proposed algorithm are close to the representative samples and far away from the vague samples. This can be owed to sample weighting. Sample weighting emphasizes the effect of the representative samples, and simultaneously weakens the interference of vague samples.

Besides, it can be seen that, in Table 2, the CPU processing time taken to perform the proposed algorithm is more than that of the FCM algorithm. This is because the proposed algorithm includes several methods and is more complex than the FCM algorithm. But, through adopting a cluster validity index, feature weighting, and sample weighting in the hybrid algorithm, the three shortcomings of the FCM algorithm can be overcome and the clustering accuracy can be greatly improved. There is always a trade-off between computational complexity and performance improvement. A reasonable compromise between these two competing requirements is desirable. It is true that, in the proposed algorithm, the performance improvement has the price of leaning feature weights, computing sample weights, and *PBMF*. Computing the sample weights and *PBMF* is relatively simple. Thus, the computational complexity mostly depends on the learning process of the feature weights. The diagnosis result indicates that the time (only 2.765 s) taken by the hybrid algorithm is acceptable. Thus, the price of computational complexity is worthy and the hybrid clustering algorithm can be applied to the fault diagnosis of rotating machinery effectively.

5 Conclusion

In this paper, we present a new hybrid clustering algorithm to overcome the existing three shortcomings in the fuzzy *c*-means (FCM) algorithm. Firstly, three techniques, i.e., feature weighting based on a three-layer feed forward neural network (FFNN), sample weighting using a distri-

bution density function, and a cluster validity index, are introduced. Then, they are incorporated into the FCM algorithm to create the presented algorithm. Unlike the existing improved algorithms, which only overcome one of the problems of the FCM algorithm, the hybrid clustering method overcomes all three problems simultaneously.

The presented algorithm is applied to the incipient fault diagnosis of locomotive roller bearings. The result shows that it not only automatically determines the proper number of clusters, but it also significantly improves the clustering accuracy. Thus, the proposed algorithm is a promising approach to the incipient fault diagnosis of rollers bearings. Of course, the algorithm discussed here can also be applied to the fault diagnosis of other rotating machinery.

References

1. Li Z, Wu Z, He Y, Fulei C (2005) Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery. Mech Syst Signal Proc 19 (2):329–339
2. Sun Q, Chen P, Zhang D, Xi F (2004) Pattern recognition for automatic machinery fault diagnosis. J Vib Acoust—Trans ASME 126(2):307–316
3. Zhang S, Mathew J, Ma L, Sun Y (2005) Best basis-based intelligent machine fault diagnosis. Mech Syst Signal Proc 19 (2):357–370
4. Ericsson S, Grip N, Johansson E, Persson LE, Sjöberg R, Strömberg JO (2005) Towards automatic detection of local bearing defects in rotating machines. Mech Syst Signal Proc 19 (3):509–535
5. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Proc 20(7):1483–1510
6. Staszewski WJ, Worden K, Tomlinson GR (1997) Time–frequency analysis in gearbox fault detection using the Wigner–Ville distribution and pattern recognition. J Eng Appl Sci 11(5):673–692
7. Li R, Chen J, Wu X, Alugongo AA (2005) Fault diagnosis of rotating machinery based on SVD, FCM and RST. Int J Adv Manuf Technol 27(1–2):128–135
8. Wu X, Chen J, Li R, Li F (2006) Internet-based remote monitoring and fault diagnosis system. Int J Adv Manuf Technol 28(1–2):162–175
9. Liu SC, Liu SY (2003) An efficient expert system for machine fault diagnosis. Int J Adv Manuf Technol 21(9):691–698
10. Liu Y, Shen Y, Liu Z (2000) An approach to fault diagnosis for non-linear system based on fuzzy cluster analysis. In: Proceedings of the 17th IEEE Conference on Instrumentation and Measurement Technology (IMTC 2000), Baltimore, Maryland, May 2000, vol 3, pp 1469–1473
11. Zhou L (2005) 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, Guangzhou, China, August 2005, vol 7, pp 3996–3999
12. Chen A-P, Lin C-C (2001) Fuzzy approaches for fault diagnosis of transformers. Fuzzy Sets Syst 118(1):139–151
13. Xu R, Wunsch D 2nd (2005) Survey of clustering algorithms. IEEE Trans Neural Networks 16(3):645–678
14. Liao TW (2005) Clustering of time series data—a survey. Pattern Recognit 38(11):1857–1874
15. Dunn JC (1974) Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems. J Cybernetics 4:1–15
16. Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York
17. Gao XB, Xie WX (2000) Advances in theory and applications of fuzzy clustering. Chin Sci Bull 45(11):961–970
18. Chan YL, Ching WK, Ng KP, Huang JZ (2004) An optimization algorithm for clustering using weighted dissimilarity measures. Pattern Recognit 37(5):943–952
19. Wang X, Wang Y, Wang L (2004) Improving fuzzy c-means clustering based on feature-weight learning. Pattern Recognit Lett 25(10):1123–1132
20. Frigui H, Nasraoui O (2004) Unsupervised learning of prototypes and attribute weights. Pattern Recognit 37(3):567–581
21. Sun H, Wang S, Jiang Q (2004) FCM-based model selection algorithms for determining the number of clusters. Pattern Recognit 37(10):2027–2037
22. Tsekouras GE, Sarimveis H (2004) A new approach for measuring the validity of the fuzzy c-means algorithm. Adv Eng Softw 35(8–9):567–575
23. Pakhira MK, Bandyopadhyay S, Maulik U (2005) A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification. Fuzzy Sets Syst 155(2):191–214
24. Tang Y, Sun F, Sun Z (2005) Improved validation index for fuzzy clustering. In: Proceedings of the American Control Conference (ACC 2005), Portland, Oregon, June 2005, vol 2, pp 1120–1125
25. Wu K-L, Yang M-S (2005) A cluster validity index for fuzzy clustering. Pattern Recognit Lett 26(9):1275–1291
26. Kim DW, Lee KH, Lee D (2004) On cluster validity index for estimation of the optimal number of fuzzy clusters. Pattern Recognit 37(10):2009–2025
27. Pakhira MK, Bandyopadhyay S, Maulik U (2004) Validity index for crisp and fuzzy clusters. Pattern Recognit 37(3):487–501
28. Basak J, De RK, Pal SK (1998) Unsupervised feature selection using a neuro-fuzzy approach. Pattern Recognit Lett 19(11):997–1006
29. Yeung DS, Wang XZ (2002) Improving performance of similarity-based clustering by feature weight learning. IEEE Trans Pattern Anal Mach Intell 24(4):556–561
30. Bezdek JC (1975) Mathematical models for systematic and taxonomy. In: Proceedings of the 8th International Conference on Numerical Taxonomy, Oeiras, Portugal, pp 143–166
31. Bezdek JC (1973) Cluster validity with fuzzy sets. J Cybernetics 3 (3):58–73
32. Fukuyama Y, Sugeno M (1989) A new method of choosing the number of clusters for the fuzzy c-means method. In: Proceedings of the 5th Fuzzy System Symposium, Kobe, Japan, pp 247–250
33. Davé RN (1996) Validating fuzzy partitions obtained through c-shells of circular clustering. Pattern Recognit Lett 17(6):613–623
34. Xie X, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13(8):841–847
35. Gath I, Geva AB (1989) Unsupervised optimal fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 11(7):773–781

36. Kwon SH (1998) Cluster validity index for fuzzy clustering. Electron Lett 34(22):2176–2177

37. Bezdek JC (1998) Pattern recognition. In: Handbook of fuzzy computation. IOP Publishing, New York

38. Pal NR, Bezdek JC (1995) On cluster validity for the fuzzy c-means model. IEEE Trans Fuzzy Syst 3(3):370–379

39. Peng ZK, Tse PW, Chu FL (2005) A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. Mech Syst Signal Proc 19(5):974–988

40. Wong MLD, Jack LB, Nandi AK (2006) Modified self-organising map for automated novelty detection applied to vibration signal monitoring. Mech Syst Signal Proc 20(3):593–610

41. Howard IM (1994) A review of rolling element bearing vibration—detection, diagnosis and prognosis. Defense Science and Technology Organization, Australia

42. Li CQ, Pickering CJD (1992) Robustness and sensitivity of non-dimensional amplitude parameters for diagnosis of fatigue spalling. Cond Monit Diagn Technol 2(3):81–84