Article

Vibration Sensor-Based Bearing Fault Diagnosis Using Ellipsoid-ARTMAP and Differential Evolution Algorithms

Chang Liu, Guofeng Wang *, Qinglu Xie and Yanchao Zhang

Key Laboratory of Mechanism Theory and Equipment Design of Ministry of Education, Tianjin University, Tianjin 300072, China; E-Mails: lc8542969@163.com (C.L.); xql246888@126.com (Q.X.); zhangyanchao1991@126.com (Y.Z.)

* Author to whom correspondence should be addressed; E-Mail: gfwangmail@tju.edu.cn; Tel.: +86-022-8183-1363; Fax: +86-022-2740-6260.

Abstract: Effective fault classification of rolling element bearings provides an important basis for ensuring safe operation of rotating machinery. In this paper, a novel vibration sensor-based fault diagnosis method using an Ellipsoid-ARTMAP network (EAM) and a differential evolution (DE) algorithm is proposed. The original features are firstly extracted from vibration signals based on wavelet packet decomposition. Then, a minimum-redundancy maximum-relevancy algorithm is introduced to select the most prominent features so as to decrease feature dimensions. Finally, a DE-based EAM (DE-EAM) classifier is constructed to realize the fault diagnosis. The major characteristic of EAM is that the sample distribution of each category is realized by using a hyper-ellipsoid node and smoothing operation algorithm. Therefore, it can depict the decision boundary of disperse samples accurately and effectively avoid over-fitting phenomena. To optimize EAM network parameters, the DE algorithm is presented and two objectives, including both classification accuracy and nodes number, are simultaneously introduced as the fitness functions. Meanwhile, an exponential criterion is proposed to realize final selection of the optimal parameters. To prove the effectiveness of the proposed method, the vibration signals of four types of rolling element bearings under different loads were collected. Moreover, to improve the robustness of the classifier evaluation, a two-fold cross validation scheme is adopted and the order of feature samples is randomly arranged ten times within each fold. The results show that DE-EAM classifier can recognize the fault categories of the rolling element bearings reliably and accurately.

1. Introduction

Fault diagnosis, which includes fault detection and isolation (FDI) [1,2], fault tolerant control (FTC) [3,4] and fault classification [5,6], plays an important role in automation systems, process engineering and mechanical equipment. In these research fields, fault recognition of rolling element bearings has attracted more and more attention of many researchers. As one of the most critical mechanical components, it has found wide applications in rotating machinery such as motors, pumps, compressors and wind power generators, *etc*. However, its defects may cause malfunctions or even lead to catastrophic failure of the rotating machinery if not diagnosed in time. The available studies have shown that bearing faults account for 40% of all machinery failures [7,8]. Therefore, effective fault monitoring is necessary for reducing maintenance costs and avoiding unscheduled downtime. Different from some diagnosing methods used in control systems and industrial processes, data driven models are commonly adopted to realize condition monitoring and fault diagnosis of rolling element bearings [5,6,9]. In comparison with the model driven method used in [1–4], the advantage of data-based modeling lies that the model can be built without understanding the complex internal mechanics of the mechanical system. For rolling element bearings, an accurate analytic dynamic model to depict the characteristic of rolling element bearings is hard to construct because there are so many sub-parts which include inner race, outer race, elements, *etc*. Therefore, data driven models based on historic data is preferred in real industrial applications and its robustness mainly depends on effective signal preprocessing, feature extraction and statistical modeling strategies.

In the past two decades, various kinds of pattern recognition techniques such as artificial neural networks (ANNs) [10–12] and support vector machines (SVMs) [13–15] have been successfully applied to classifying bearing faults. These methods have been proven to be capable of obtaining favorable classification accuracy in some research areas [5]. However, in real applications, the useful information is usually contaminated by external noise or internal load variation. In such cases, the classifiers based on the above methods are liable to cause over-fitting phenomena and deteriorate the classification performance [6]. At the same time, these methods seek to depict the boundary of the classifier as the function of several support vectors or network weight values, which makes it hard to characterize the complex decision boundary if the spatial distributions of the training samples is disperse and irregular.

In this paper, Ellipsoid-ARTMAP (EAM), which is based on adaptive resonance theory (ART) [16], is presented to realize the task of classifying complex samples due to its strong local and distributed representation ability [17–19]. Unlike Radius Basis Function (RBF) neural network, which is utilized to estimate the probability density of the data streams and judge the occurrence of the fault by comparing the probability density of the current data with the previous one [20], EAM is mainly proposed to depict the spatial distribution of the sample data for each category. One major characteristic of EAM is that distributed hyper-ellipsoid clusters are utilized to realize the geometric representation of each

category. Therefore, the decision boundary of the dispersive samples can be depicted accurately and flexibly. Another is that smoothing operation is performed during the updating process of these ellipsoid nodes, which makes it insensitive to noise and can avoid over-fitting. Xu *et al*. [21] applied the EAM network to classify the gene expression data generated by DNA microarray experiments. They claimed that a good performance can be achieved. Anagnostopoulos *et al*. [22,23] utilized EAM for the Circle-in-a-Square classification problem. The authors proved that EAM demonstrated better ability in clustering and classification tasks than other ARTMAP networks. However, it can be also demonstrated that the performance of EAM network is greatly affected by its parameter selection strategy. In fact, there are four important parameters that need to be determined a *priori*: vigilance parameter, learning rate, effective diameter and ratio of minor-to-major axes lengths. Each parameter exerts a significant impact on the performance of the EAM classifier. Currently, the selection of these parameters mainly depends on empirical selection or trial and error methods, which inevitably bring inaccuracy and uncertainty to the classification performance. To solve this problem, a differential evolution (DE) algorithm is adopted in this paper to optimize the network parameters of EAM. One reason for selecting DE is that it has strong global optimization ability and imposes no demands for the analytical mathematical expression of the fitness function, which is paramount for EAM parameter optimization because such an equation does not exist during the classification process. In addition, the DE has several advantages: compared with traditional evolutionary optimization algorithms, its fewer control parameters make DE simpler and more straightforward to implement. Meanwhile, the differential operation of the candidate parameter vectors optimization makes it more effective to find a global optimum when handling large scale and multi-objective optimization problems [24]. Up to now, DE algorithm has been used in many applications such as system design [25], partitioned clustering [26] and classification tasks [27].

Based on the DE optimization algorithm and EAM classifier, a vibration sensor-based fault diagnosis system is constructed. In this framework, vibration signals were collected and wavelet packet decomposition (WPD) is introduced to extract features to characterize the fault related information. Meanwhile, the minimum-redundancy maximum-relevancy (mRMR) feature selection method is employed to decrease the redundancy and irrelevancy of these features. Finally, the EAM classifier is constructed based on the selected salient features to realize fault classification of rolling element bearings and DE is integrated with EAM to optimize the classifier parameters during the training process. In order to get satisfactory accuracy without lowering the recognition speed, both the classification accuracy and number of nodes are adopted as the fitness functions of the DE algorithm. Meanwhile, a new exponential selection criterion is presented to balance the influence of both functions and choose the final parameters. To verify the effectiveness of the DE-EAM classifier, four kinds of faulty vibration signals (normal, inner race fault, outer race fault and ball fault) were collected from a test rig. Meanwhile, two-fold cross validation and random order strategies are adopted to evaluate the performance of classifier robustly and accurately. The final results of the bearing faults diagnosis show that DE-EAM method can accurately recognize the fault categories.

The rest of the paper is organized as follows: in Section 2, the principles of the EAM classifier and DE optimization algorithm are explained, respectively. In Section 3, the framework of the DE-EAM- based monitoring system is presented. Moreover, the principle of WPD feature extraction and mRMR feature selection are discussed in detail. In Section 4, the effectiveness of the proposed method is verified by

classification of four kinds of bearing faults under a cross validation and random order strategy. Some useful conclusions are drawn in Section 5.

2. Principle of the DE-EAM Algorithm

2.1. Ellipsoid-ARTMAP (EAM)

2.1.1. Network Structure

Figure 1 shows the framework of EAM network. It can be seen that a typical EAM structure consists of three layers: input layer F_1, representation layer F_2 and mapping layer F_3. F_1 and F_2 layer are connected by the template vector $\mathbf{w}_j$, which is utilized to encode the input sample into the *j*th node in F_2 layer. If the node passed the vigilance test, the commitment test is then applied on these nodes and the winner nodes are mapped into the F_3 layer so as to get the classification results. During the training process, a match tracking (MT) process [16] will be invoked when the output of F_3 layer is incorrect.

Figure 1. The framework of the Ellipsoid-ARTMAP network.
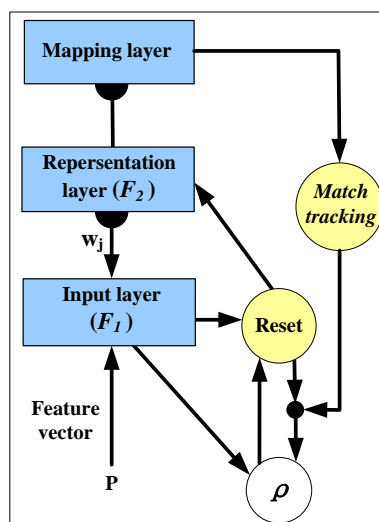
2.1.2. Node Representation

Figure 2 shows a two-dimensional representation of an EAM node. Each node *j* is described by a template vector $\mathbf{w}_j = [\mathbf{m}_j, \mathbf{d}_j, R_j]$ where $\mathbf{m}_j$ is the center of hyper-ellipsoid, $\mathbf{d}_j$ is the direction vector of node, which coincides with the direction of the hyper-ellipsoid's major axis, and R_j is called the radius of the node, which equals half the length of the major axis.

Figure 2. Two-dimensional representation of an EAM node.
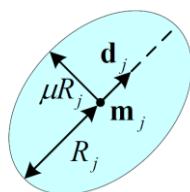
2.1.3. Distance Calculation

The distance between an input sample **x** and the node *j* is given by [16]:

$$dis(\mathbf{x}|\mathbf{w}_j) = \max\left\{\left\|\mathbf{x} - \mathbf{m}_j\right\|_{C_j}, R_j\right\} - R_j \tag{1}$$

where $\left\|\mathbf{x} - \mathbf{m}_j\right\|_{C_j}$ is the distance between a sample **x** and the center $\mathbf{m}_j$ of the node *j*, which can be expressed as [22]:

$$\left\|\mathbf{x} - \mathbf{m}_j\right\|_{C_j} = \begin{cases} \dfrac{1}{\mu}\sqrt{\left\|\mathbf{x} - \mathbf{m}_j\right\|_2^2 - (1 - \mu^2)\left[d_j^T(\mathbf{x} - \mathbf{m}_j)\right]^2} & if \quad \mathbf{d}_j \neq 0 \\ \left\|\mathbf{x} - \mathbf{m}_j\right\|_2 & if \quad \mathbf{d}_j = 0 \end{cases} \tag{2}$$

where $\left\|\cdot\right\|_2$ means the usual Euclidian (L2) norm, μ is the ratio of minor-to-major axes lengths and C_j is node shape matrix. As shown in Figure 2, the shaded areas denote the representation regions of the node *j*, which are the set of points that satisfy the following condition:

$$dis(\mathbf{x}|\mathbf{w}_j) = 0 \quad \Rightarrow \quad \left\|\mathbf{x} - \mathbf{m}_j\right\|_{C_j} \leq R_j \tag{3}$$

2.1.4. Vigilance Test (VT) and Commitment Test (CT)

For an input sample **x**, the node *j* features two important values which are called the category match function (CMF) value $\rho^*(\mathbf{w}_j|\mathbf{x})$ and the category choice function (CCF) value $T^*(\mathbf{w}_j|\mathbf{x})$ [16]:

$$\rho^*(\mathbf{w}_j|\mathbf{x}) = 1 - \frac{R_j + \max\left\{\left\|\mathbf{x} - \mathbf{m}_j\right\|_{C_j}, R_j\right\}}{D_m} \tag{4}$$

$$T^*(\mathbf{w}_j|\mathbf{x}) = \frac{D - R_j - \max\left\{\left\|\mathbf{x} - \mathbf{m}_j\right\|_{C_j}, R_j\right\}}{D_m - 2 * R_j + a} \tag{5}$$

where D_m is called effective diameter, *a* is the choice parameter. The choice parameter *a* is a really small positive value which has no obvious influence on EAM performance. In this paper, *a* is selected as 0.001. Usually, uncommitted nodes have a constant CCF value $T_u = D_m /(2D_m w + a)$. The parameter ω is usually chosen as $\omega \geq 0.5$ to ensure the stability of EAM. If the CMF value and CCF value of the node *j* for the given input sample **x** are larger than initial vigilance parameter $\bar{\rho}$ and T_u, this node is called a committed node. These two comparison processes are called the vigilance test (VT) and commitment test (CT), respectively.

2.1.5. Node Updating

During the training process, the committed node is updated according to the input sample. That is, its template vector $\mathbf{w}_j$ is recalculated according to the formula below [23]:

$$R_j^{new} = R_j^{old} + \frac{\gamma}{2}(\max\left\{R_j^{old}, \left\|\mathbf{x} - \mathbf{m}_j^{old}\right\|_{C_j^{old}}\right\} - R_j^{old}) \tag{6}$$

$$\mathbf{m}_j^{new} = \mathbf{m}_j^{old} + \frac{\gamma}{2}\left(1 - \frac{\max\{R_j^{old}, \left\|\mathbf{x} - \mathbf{m}_j^{old}\right\|_{C_j^{old}}\}}{\left\|\mathbf{x} - \mathbf{m}_j^{old}\right\|_{C_j^{old}}}\right)(\mathbf{x} - \mathbf{m}_j^{old}) \tag{7}$$

$$\mathbf{d}_j = \frac{\mathbf{x}_{(2)} - \mathbf{m}_j}{\left\|\mathbf{x}_{(2)} - \mathbf{m}_j\right\|_2} \tag{8}$$

where $\gamma \in (0,1]$ denotes the learning rate. If γ is selected as 1, it means that the EAM has the ability of fast learning. $\mathbf{x}_{(2)}$ represents the second sample encoded by the node j.

A two-dimensional description of node updating based on Equations (6)–(8) is shown in Figure 3. It can be shown that the EAM node can only grow in size and never be destroyed during the training process. The node's new representation region is a minimum hyper-ellipsoid which contains both the old region and the new sample to be encoded. In addition, when the EAM node's direction vector is set, it will not change during the node updating process. Besides that, due to the minimum hyper-ellipsoid is used to characteristic the old region and the new pattern, the E^{new} and E^{old} can only touch at one point.

**Figure 3.** The updating process of a two-dimensional EAM node.
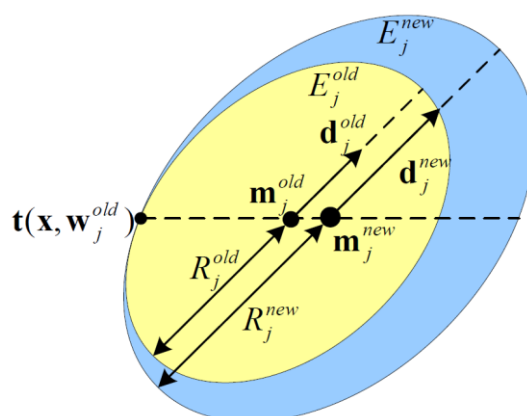
2.1.6. Training and Classification

The complete training process of the EAM classifier is given in the following steps:

Step 1: The network parameters are initialized.

Step 2: The distance calculation is started based on Equation (3) to judge whether the training sample belongs to the representation region of the node j or not. If not, the input sample will undergo the VT. That is to say, the value of $\rho^*(\mathbf{w}_j|\mathbf{x})$ larger than $\bar{\rho}$ will be selected into a candidate set **S**, otherwise, it will be removed.

Step 3: CCF value of all member nodes in **S** is calculated and compared with T_u. The nodes that cannot pass the CT will be removed from **S**. If no nodes pass the VT and CT (here, **S** is empty), new node will be created and initiated.

Step 4:   If the set **S** is non-empty after the VT and CT, the labels of the training sample and the chosen node *j* will be used for comparison. If the selected node *j* shares the same label with the input sample, the template vector $\mathbf{w}_j$ is updated. If the label is not correct, the match tracking (MT) process will be activated. The flowchart of the EAM training is given in Figure 4.

Figure 4. Flowchart of EAM training.
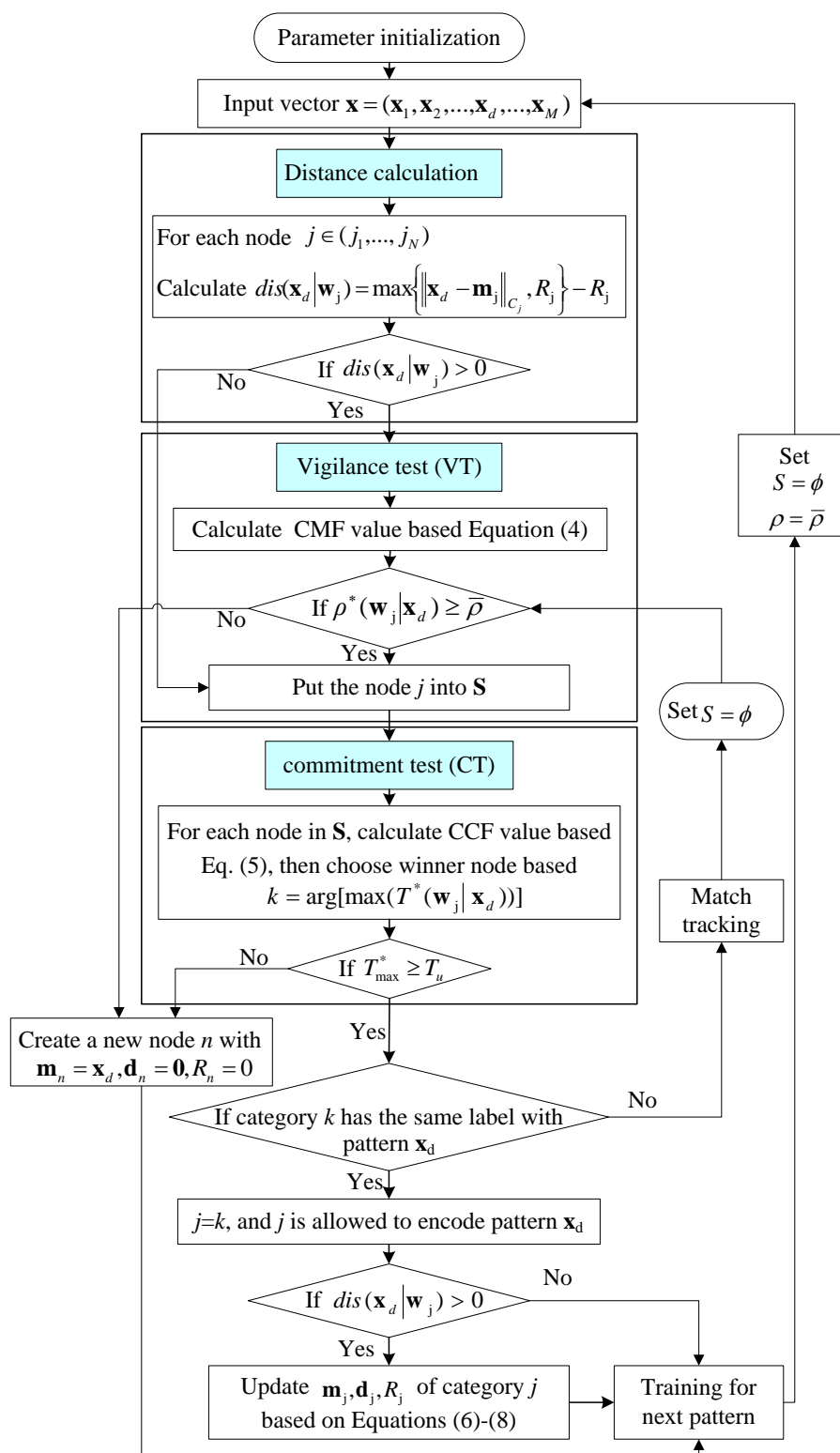
The EAM classification is similar to the training process. The only difference is that the MT process is neglected if the EAM is used for classification. If all nodes are not committed, the output of the EAM classifier is −1, which denotes that an abnormal result appears.

As mentioned above, the following four parameters have an important influence on the performance of the EAM classifier: the effective diameter D_m, the ratio of minor-to-major axes lengths $\mu \in (0,1]$, the vigilance parameter $\bar{\rho} \in [0,1]$ and the learning rate $\gamma \in (0,1]$. Vigilance parameter $\bar{\rho}$ is a threshold value which is a necessary condition for judging whether the input belongs to the corresponding node or not. Learning rate γ assure that EAM has a flexible learning ability range from slow learning to fast learning. Effective diameter D_m is always chosen to ensure the stability of EAM. Ratio of minor-to-major axes lengths μ determines the shape of the hyper-ellipsoid node. To improve the accuracy of the classifier, differential evolution (DE) algorithm is introduced to obtain the optimal combination of these parameters. The principle of DE is depicted in the following section.

2.2. Differential Evolution (DE)

As a member of the evolutionary optimization algorithm class, the DE algorithm uses three classical operators: mutation, crossover and selection, to generate trial parameter vectors and select the final solution with best fitness [28]. The main advantage of evolutionary algorithms is their global search abilities with no risk of falling into the local minimum region. In contrast, the gradient-based local research algorithm is not suitable for optimizing the EAM network structure [29,30] because the analytical mathematical expression of the objective function is not available. Moreover, although there are many variants, the traditional DE/rand/1/bin algorithm is still adopted in this paper because of its maturity, robustness and fast convergence features [31–33].

Each generation in DE comprises N_p populations. The population is expressed as $\mathbf{Y}_{i,g}$. $i \in (1, ..., N_p)$ is the index of population and g represents the number of generations. Each population contains parameters that need to be optimized. Here in this paper, four parameters of the EAM network need to be optimized, that is to say, $\mathbf{Y}_{i,g} = \{D_m, \mu, \bar{\rho}, \gamma\}$. The main steps of DE algorithm are given as follows:

Step 1: Initialization

Set the boundaries of the optimized parameters and initialize their value in the first generation according to a uniform probability distribution.

Step 2: Mutation

Mutation operation is to add a vector differential to a population vector of individuals. For each vector $\mathbf{Y}_{i,g}$, the mutant vector is formed based on the following rule:

$$\mathbf{V}_{i,g+1} = \mathbf{Y}_{r1,g} + F(\mathbf{Y}_{r2,g} - \mathbf{Y}_{r3,g}) \tag{9}$$

where, $\mathbf{Y}_{r1,g}$, $\mathbf{Y}_{r2,g}$ and $\mathbf{Y}_{r3,g}$ are selected from the populations randomly. The indices $r1$, $r2$ and $r3$ are different from each other and also differ with the current individual i (*i.e.*, $r1 \neq r2 \neq r3 \neq i$). $F \in [0,1]$ is a scaling factor which controls the amplitude of the difference vector $(\mathbf{Y}_{r2,g} - \mathbf{Y}_{r3,g})$.

Step 3: Crossover

After the mutation operation, DE utilizes crossover operation to build the trial vectors. Specially, the elements from the parent vector $\mathbf{Y}_{j,i,g}$ will be combined with $\mathbf{V}_{j,i,g}$ so as to produce the new element $\mathbf{U}_{j,i,g+1}$, which is defined as:

$$\mathbf{U}_{j,i,g+1} = \begin{cases} \mathbf{V}_{j,i,g+1} & if\ (rand_j < C)\ or\ (\tau = j) \\ \mathbf{Y}_{j,i,g} & otherwise \end{cases} \tag{10}$$

where, $j = 1,2,\ldots V$, $rand_j \in [0,1]$ is a random number, $C \in [0,1]$ is a predefined crossover parameter, $\tau \in (1,2,\ldots,V)$ is a randomly selected index. $\tau = j$ enables at least one of the parameters in the offspring to be different from their parent.

Step 4: Selection

Selection operation is utilized to choose the better offspring based on the follow equation:

$$\mathbf{Y}_{i,g+1} = \begin{cases} \mathbf{U}_{i,g+1} & if\ f(\mathbf{U}_{i,g+1}) \leq f(\mathbf{Y}_{i,g}) \\ \mathbf{Y}_{i,g} & otherwise \end{cases} \tag{11}$$

where, $f(*)$ denotes the fitness function. It can be seen that the solution vector with better fitness value is chosen as the new generation. Obviously, the selection process guarantees that the population fitness is either improving or at least maintaining the best values so far.

Step 5: Termination

The process of the mutation, crossover and selection is repeated until the maximum generation number g_{max} is satisfied. The selection of g_{max} depends on the trend of the fitness value fluctuation. Generally, the fitness function tends to be stable with the proceeding of the iteration process. The flowchart of DE is given in Figure 5.

**Figure 5.** Flowchart of DE iterative optimization.
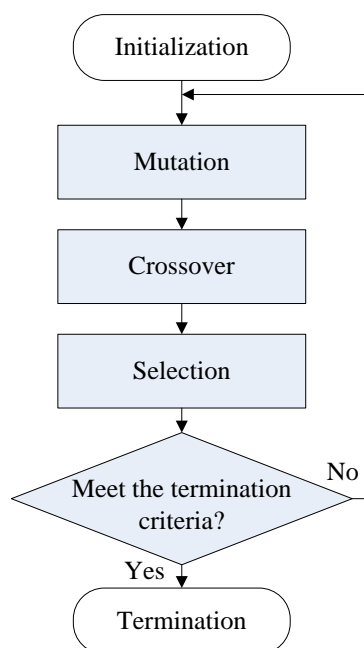
3. Diagnosis System Based on DE-EAM

3.1. Framework

The framework of the DE-EAM based fault diagnosis system is shown in Figure 6. It can be seen that the whole system is composed of four parts: (1) data collection; (2) feature extraction; (3) feature selection; (4) fault classification. Vibration signals from different bearing fault types are collected and Wavelet Packet Decomposition (WPD) is adopted to extract the characteristic features. Then, mRMR method is used to decrease the redundancy and irrelevancy of the original features. Finally, the EAM classifier is constructed to perform bearing fault classification and DE algorithm is used to optimize the network parameters of the EAM classification. The detailed description of each part is given in the sections that follow.

Figure 6. Architecture of bearing fault diagnosis system.
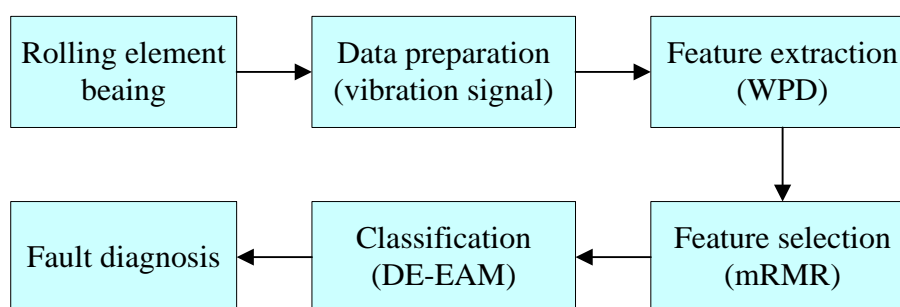
3.2. Data Collection

Vibration analysis, due to its simplicity and effectiveness, has been wildly used in bearing fault diagnosis [34–36]. Konar and Chattopadhyay [14] have claimed that the vibration response is the most reliable technique to detect and diagnose the localized defects. In this paper, vibration signals were collected and used to depict the dynamic characteristic of rolling element bearings. As well known, the vibration signal contains both the low-frequency impact and high-frequency structural information. Therefore, the WPD algorithm is adopted in the following section to extract the complete characteristics of the rolling element bearings.

3.3. Feature Extraction

WPD is a kind of multi-resolution signal processing algorithm. The main characteristic of WPD is that both the low-frequency and high-frequency information are decomposed simultaneously. The decomposition structure of WPD is shown in Figure 7. In this paper, the input data d denotes the vibration sensory signal collected and its length is N. The corresponding wavelet coefficients of d in the jth level for WPD can be written as [37]:

$$d_j^n(k) = \sum_m h_0(m-2k)d_{j-1}^{n/2}(m) \qquad (n \ is \ even) \tag{12}$$

$$d_j^n(k) = \sum_m h_1(m-2k)d_{j-1}^{(n-1)/2}(m) \qquad (n \ is \ odd) \tag{13}$$

where, n is the index of sub-bands, h_0 and h_1 are a pair of orthogonal filters for decomposition, h_0 is low pass filter and h_1 is a high pass filter. By decomposing the signal into different frequency band, the influence of noisy disturbance will be weakened obviously. Therefore, component of useful signal at each level can be enhanced correspondingly, which improves the robustness of the extracted features. Here, the sub-band energy E_j^n at jth level are then selected as initial features whose mathematical expression can be calculated as:

$$E_j^n = \sum_k (d_j^n(k))^2$$

(14)

As described above, when the decomposed level of WPD is selected to be u, the number of the feature vectors is 2^u. The energy feature in each sub-band is calculated according to Equation (12) and the feature set F is generated correspondingly as $F = \left\{ E_u^1, E_u^2, ..., E_u^{2^u} \right\}$.

Figure 7. High and low frequency decomposition of vibration signals based on WPD.
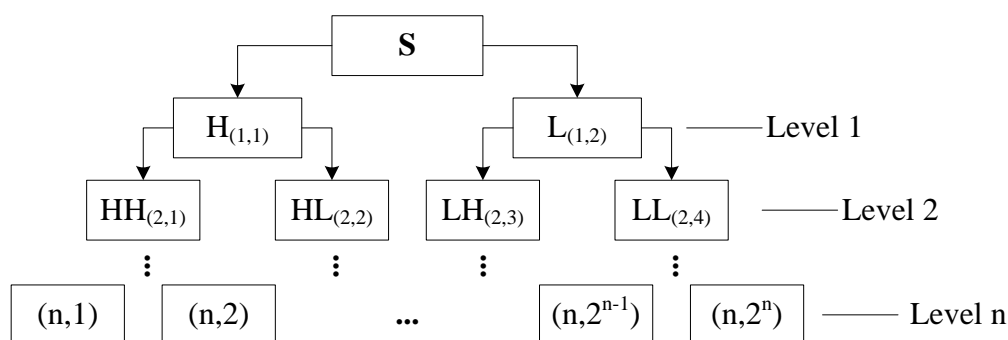
3.4. Feature Selection

Irrelevant and redundant features negatively affect the classification performance of the classifier. In this study, a sequential forward selection algorithm named the minimum-redundancy maximum-relevance (mRMR) [38] is employed to search for the optimal feature combination from the original features. The mRMR algorithm utilizes mutual information (MI) to select those optimal features which can best fulfill the minimal redundancy (Min-Redundancy) and maximal relevance (Max-Relevance) criterion.

Within the mRMR algorithm, MI is to estimate the correlation between two variables so as to quantify both the relevance and redundancy. The mathematical expression of MI is given as:

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \, dxdy$$

(15)

where x and y are two vectors, $p(x, y)$ is their joint probabilistic density, both $p(x)$ and $p(y)$ denote the marginal probabilistic densities.

F is defined as the original feature set which concludes h features and class label c. Let F_a denote the already selected feature set with m features, F_b denotes the candidate feature set with n features. m plus n is equal to h. Initially, the relevance of the features in F_b set is calculated by:

$$D = I(f, c)$$

(16)

Then, the redundancy R of the feature f in F_b with the features f_j in F_a can be calculated by [39]:

$$R = \frac{1}{m} \sum_{f_j \in F_a} I(f, f_j) \ (j = 1, ..., m) \tag{17}$$

Finally, the maximum relevance and minimum redundancy feature in F_b are selected based on the following rule [39]:

$$\max_{f \in F_b} \left[I(f, c) - \frac{1}{m} \sum_{f_j \in F_a} I(f, f_j) \right] \tag{18}$$

Based on the mRMR criteria, the selected feature is removed from F_b and put into F_a. The new calculation process starts over based on the updated F_b and F_a. Initially, F_a is an empty set $(i.e., m = 0)$, F_b contains h features $(i.e., n = h)$. The relevance of the features in F_b set is calculated based on Equation (14) and the feature corresponding to the maximum relevance is selected as the first feature and put into F_a set. This feature evaluation process will continue $h - 1$ rounds and a rearranged features set F' based on mRMR algorithm is obtained as $F' = \{f_1', f_2', ...f_k'..., f_h'\}$. The first k features are then selected and form the final feature set $D = \{f_1', f_2', ...f_k'\}$ which are put into DE-EAM based classifier to realize the multi-fault recognition of the rolling element bearing.

4. Experimentation and Validation

4.1. Data Preparation

To verify the effectiveness of the DE-EAM-based diagnosis method, experimental data of four types of bearing states (normal bearing, inner race fault, ball fault and outer race fault) under different operating loads (1Hp, 2Hp, 3Hp) were collected. The experiment setup of test rig is shown in Figure 8. The deep groove ball bearing (type: 6205-2RS JEM SKF) was utilized to support the motor shaft and some parameters are given as: inside diameter: 0.9843 in; outside diameter: 2.0472 in; ball diameter: 0.3126 in; pitch diameter: 1.537 in. An accelerometer was mounted at the driven end of the motor to collect vibration signals. These different faults were introduced using the electric discharge machining (EDM) method. The sampling frequency was set to be 12,000 Hz because the vibration sensory signal of the rolling bearing is mainly focused on this band.

Figure 8. Experiment setup: (**a**) Picture of test rig; (**b**) Framework of test.
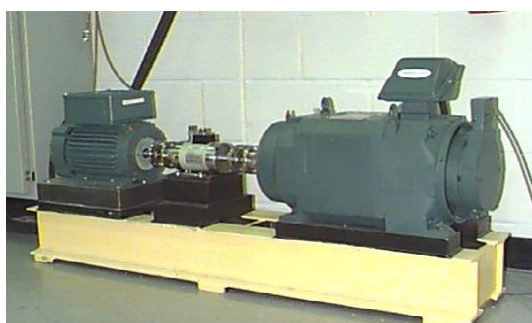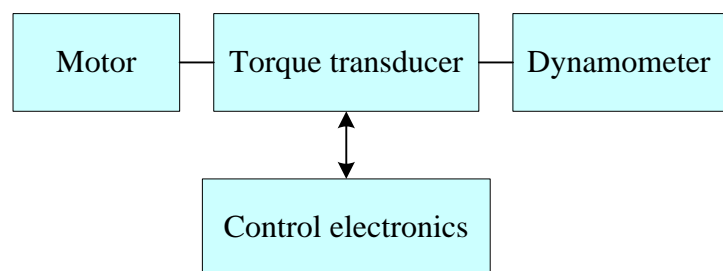
(**a**)   (**b**)

Each type of bearing was tested under three different loads (1Hp, 2Hp, 3Hp) and 100 samples were collected, respectively. The length of each sample was selected as 2048 which cover more than five

cycles of shaft rotating periods. Finally, the number of samples corresponding to each bearing state is 300 and the total number is 1200. Figure 9 shows the time domain waveforms of the four types of rolling bearing under three different loads.

Figure 9. Time domain waveform of fault bearing vibration signal under different loads. (**a**)–(**d**) denote normal bearing, inner race fault, ball fault and outer fault, respectively. The subscript 1, 2 and 3 represent the 1Hp, 2Hp, 3Hp load, respectively.
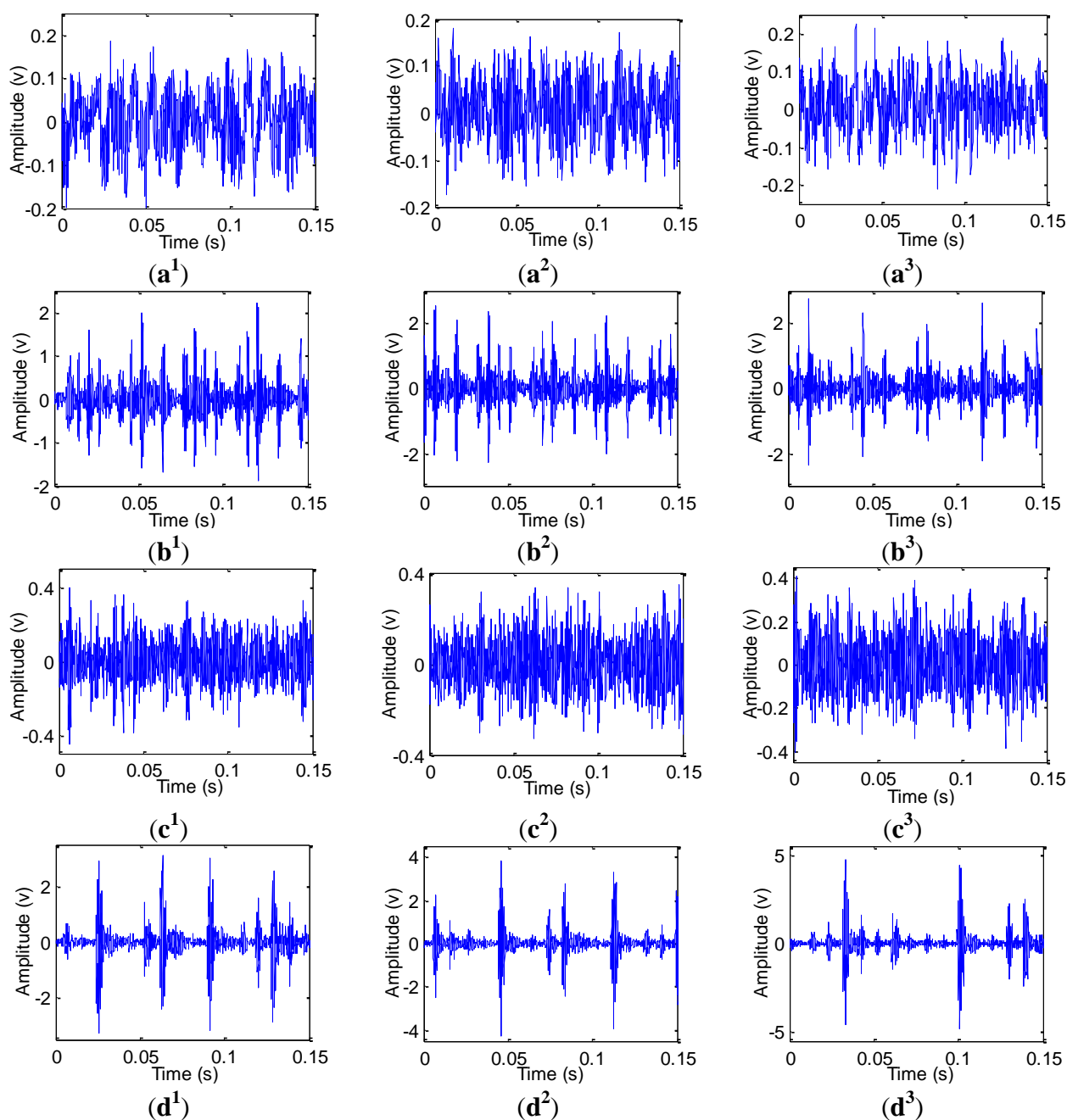

The WPD is introduced to extract features from the original vibration samples. Here, the decomposed level u is selected as 4, which lead to that the dimensionality of the original feature being 16 and a feature set F is formed with size 1200×16. The mRMR method is then applied to search a salient feature. Here, the target dimension k is chosen as 6. After mRMR based feature selection, the first six features $\{E_4^1, E_4^4, E_4^{13}, E_4^5, E_4^8, E_4^{14}\}$ are selected sequentially from F and constructed as the new

feature subset D whose size is 1200×6. Figure 10 shows the features spatial distribution of four kinds of states of rolling element bearings. It can be clearly seen that some feature data of different fault states overlap with each other, which increases the difficulty of the decision boundary representation. In addition, Figure 11 demonstrates the influence of load on the features' spatial distribution for every bearing state. It can be seen that the distribution scopes of the features under different loads are obviously different from each other, although their fault state is the same. In such a case, this casts more difficulty on depicting the spatial distribution of each class and judging the bearing states accurately.

Figure 10. Features spatial distribution of four kinds of bearing states.

(**a**)

(**b**)
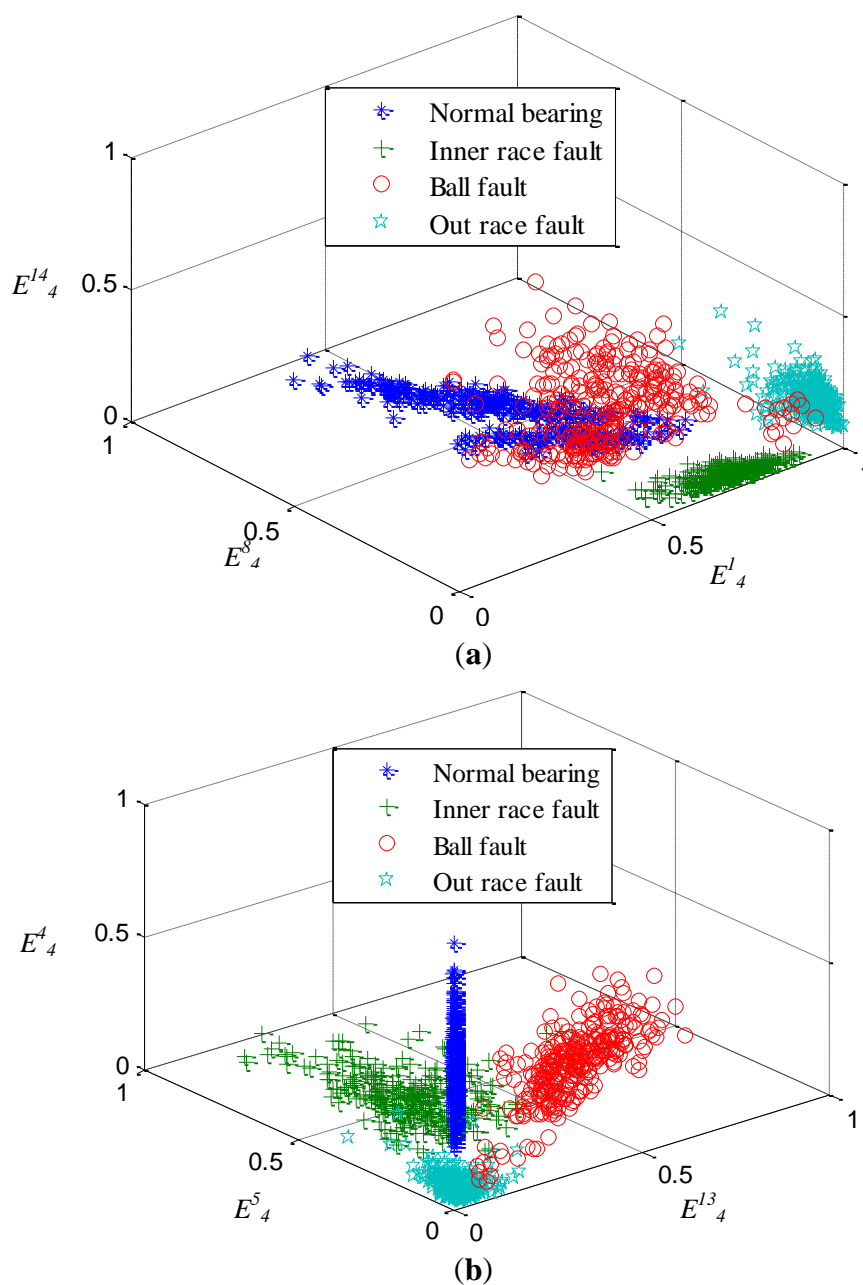
Figure 11. Features spatial distribution for each bearing state under three different loads: (**a**) normal; (**b**) inner race fault; (**c**) ball fault; (**d**) outer race fault.

(**a**)   (**b**)

(**c**)   (**d**)
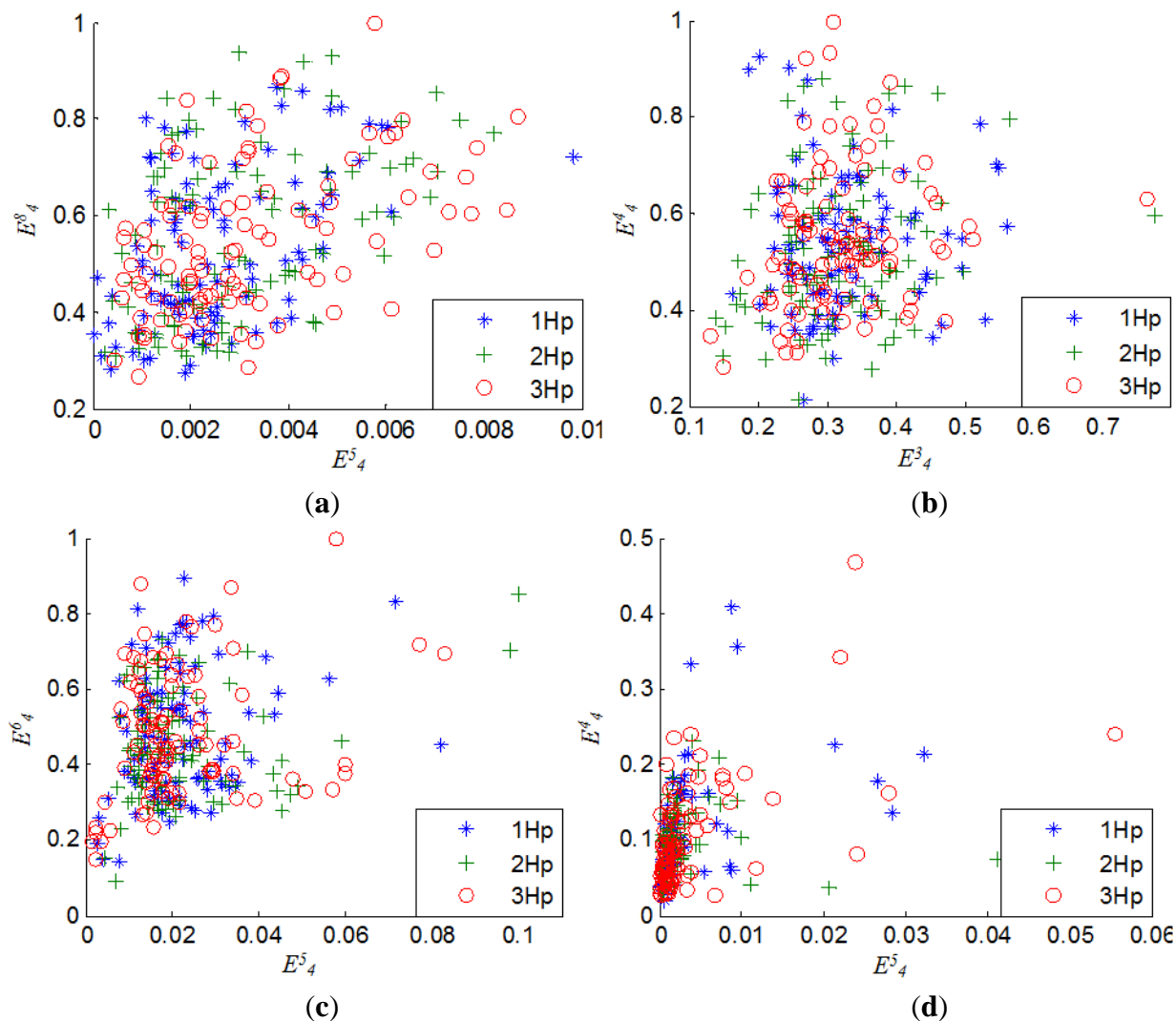
4.2. Cross-Validation

Cross-validation is a widely used scheme to evaluate the performance of a classifier. For *v*-fold cross-validation [40], all samples are randomly divided into *v* subsets with equal size and *v* iterations are performed according to the following rule: one different subset is used for validation while the remaining *v*−1 subsets are used for training to construct the classifier model. After *v* iterations, all results are averaged to produce the final estimation of the classifier. The major advantage of cross-validation is that all data are used for both training and validation, and each data is used for validation exactly once. In such case, the accuracy estimation owns a lower variance compared to an accuracy estimate using only one training and validation set. In this paper, 2-fold cross-validation is adopted to separate training samples *D* and evaluate the performance of the classifier.

4.3. Parameters Optimization

For the EAM classifier, four network parameters, vigilance parameter $\bar{\rho}$, learning rate γ, effective diameter D_m, and ratio of minor-to-major axes lengths μ need to be optimized using the DE

optimization algorithm. The control parameters in DE are defined as follows: scaling factor $F = 0.5$, crossover rate $C = 0.2$ and the number of population $N_p = 40$. In addition, the maximum generation number g_{max} is selected as 20.

The classification error (*Ce*) and relative number of nodes ($R_{nod} = nd / M$, where *nd* is the number of EAM nodes, *M* is the dimensionality of samples) are the two most important indicators to evaluate the performance of the EAM classifier. *Ce* directly denotes the classifying ability, the smaller the error, the higher the ability. *nd* represents the computing speed, the less the nodes, the faster the network. In this paper, both *Ce* and R_{nod} are selected as the fitness function of DE optimization. Figure 12 shows the variation of the two fitness functions with the generation number. It can be seen that both fitness curves tend to be smooth after 12 generations. Therefore, the selection of g_{max} is suitable for guaranteeing the convergence of the DE algorithm.

Figure 12. Variation of the fitness functions with the generation number.
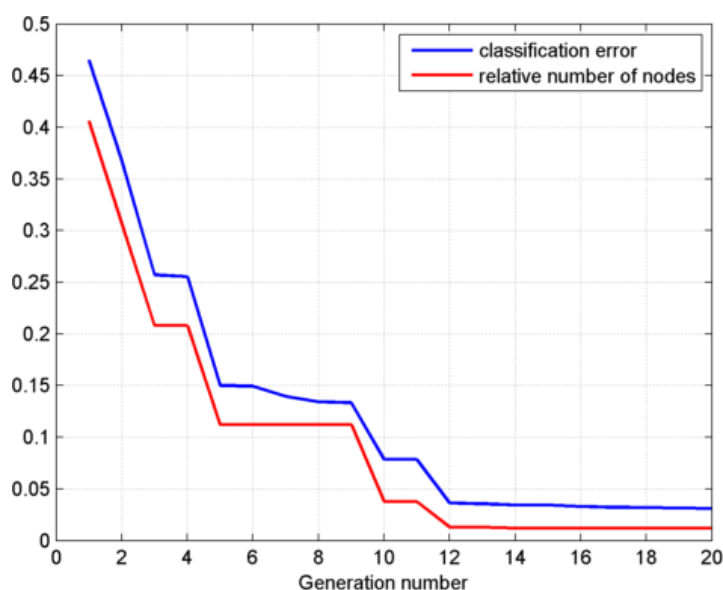
After the maximum generation is satisfied, 40 populations are obtained in which each contains four EAM parameters. To get the final solution, a compound indicator Φ_t is calculated for every possible solution and the final index *i* of the optimized parameters correspond to the minimum Φ_t value, that is:

$$\Phi_i = \min \left[e^{Ce(t)} + e^{R_{nod}(t)} \right] (t \in (1,...,N_p)) \tag{19}$$

where $Ce(t)$ and $R_{nod}(t)$ represent the classification error and relative number of nodes in *t*th population of the objective generation, respectively. After the index *i* is determined based on Equation (17), the final solution vector $\mathbf{Y}_{i,g}$ which contains four optimized parameters will be utilized to construct the EAM classifier.

4.4. Classification and Analysis

To evaluate the proposed DE-EAM classifier accurately, the order of feature samples is randomly rearranged ten times and a two-fold cross validation method is adopted simultaneously. Figure 13 shows the averaging classification accuracy of the four types of bearing state under every case. In this

figure, the *x* axis denotes the repeat times and the *y* axis represents the classification accuracy. It can be obviously seen that the minimum accuracy is 94.75% and the maximum accuracy is 97.5%. The overall averaging classification accuracy can reach 96.1%.

Figure 13. The averaging accuracy of the EAM classifier under different folds and rearrangement orders.
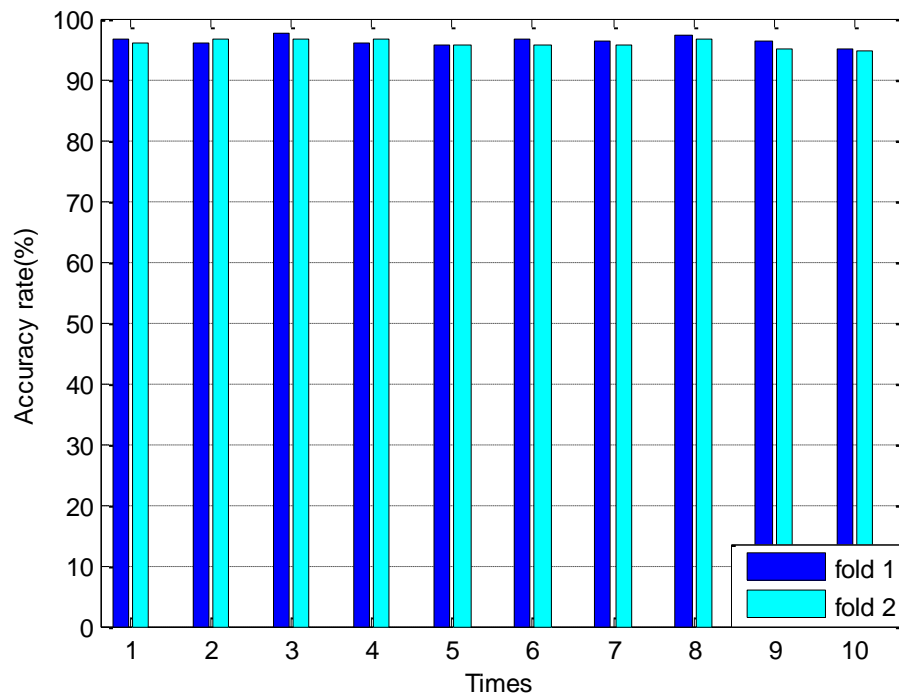
To further show the capability of the DE-EAM classifier, the accuracy of recognizing each kind of fault state is also calculated and listed in Table 1. It can be seen that for the normal bearing, the minimum accuracy is 98% and the maximum is 100%. For the recognizing of the ball pit fault, the lowest accuracy appears, although it can be seen that the average accuracy can still reach 92.9%. Therefore, it can be proven that DE-EAM shows excellent performance for recognizing the fault type of the rolling element bearings.

Table 1. The classification accuracy of DE-EAM classifier for each bearing fault state.

Fault Type	Fold Number	Times of Random Ranking										Average Value
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
Normal	1	1	1	1	1	1	1	1	1	1	1	**99.8**
	2	1	1	1	1	0.99	1	1	1	0.99	0.98	
Inner	1	0.94	0.92	0.98	0.92	0.95	0.94	0.97	0.97	0.93	0.92	**93.95**
	2	0.92	0.94	0.96	0.95	0.93	0.90	0.95	0.97	0.93	0.90	
Ball	1	0.94	0.94	0.94	0.94	0.90	0.94	0.90	0.94	0.94	0.90	**92.9**
	2	0.94	0.94	0.93	0.94	0.96	0.94	0.90	0.92	0.90	0.93	
Outer	1	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	**97.8**
	2	0.98	0.98	0.98	0.98	0.95	0.98	0.98	0.97	0.98	0.98	

5. Conclusions

In this paper, a novel monitoring system combining an EAM network with a DE optimization algorithm is proposed to realize fault diagnosis of rolling element bearings. Within this framework, the original features are firstly extracted from the vibration sensory signals using WPD to characterize the low frequency and high frequency characteristics of the rolling element bearing. Moreover, the mRMR feature selection method is utilized to reduce the redundancy and irrelevancy of the features vectors. Based on the optimal selected features, the DE-EAM method is proposed to establish the corresponding classifier to realize the diagnosis of the rolling element bearings. The main advantage of the EAM classifier lies in that hyper-ellipsoids are adopted to represent the geometric shape of the feature spatial distribution and the smoothing operation algorithm is used to get a more accurate decision boundary. Moreover, the DE algorithm is integrated with EAM to obtain the optimal network parameters by the limited evolution operations. To verify the robustness and accuracy of the DE-EAM-based system, the samples under four kinds of bearing fault status are collected and organized by random order arrangement and two-fold cross validation strategy. Moreover, the EAM classifier is constructed based on the optimal network parameters which are calculated by the DE optimization. The classification results of four kinds of rolling element bearings show that the overall average accuracy is 96.1% and the maximum average accuracy for recognizing a single fault state can reach 99.8%, which demonstrates that the DE-EAM method can accurately realize fault diagnosis of rolling element bearings.

Author Contributions

The original ideal was proposed by Chang Liu and Guofeng Wang. The further data analysis was realized by Chang Liu, Qinglu Xie and Yanchao Zhang under the supervision of Guofeng Wang. Chang Liu wrote the manuscript, and all the authors participated in amending the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Simani, S. Residual generator fuzzy identification for automotive diesel engine fault diagnosis. *Int. J. Appl. Math. Comput. Sci*. **2013**, *23*, 419–438.
2. Smarsly, K.; Law, K.H. Decentralized fault detection and isolation in wireless structural health monitoring systems using analytical redundancy. *Adv. Eng. Softw*. **2014**, *73*, 1–10.

3. Puig, V. Fault diagnosis and fault tolerant control using set-membership approaches: Application to real case studies. *Int. J. Appl. Math. Comput. Sci.* **2010**, *20*, 619–635.

4. Niemann, H.H. A model-based approach to fault-tolerant control. *Int. J. Appl. Math. Comput. Sci.* **2012**, *22*, 67–86.

5. Widodo, A.; Yang, B.-S. Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2560–2574.

6. Xu, Z.B.; Xuan, J.P.; Shi, T.L.; Wu, B.; Hu, Y.M. A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique. *Expert Syst. Appl.* **2009**, *36*, 11801–11807.

7. Tandon, N.; Choudhury, A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribol. Int.* **1999**, *32*, 469–480.

8. Zhang, P.; Du, Y.; Habetler, T.G.; Lu, B. A survey of condition monitoring and protection methods for medium voltage induction motors. *IEEE Trans. Ind. Appl.* **2011**, *47*, 34–46.

9. Patil, M.S.; Mathew, J.; Rajendrakumar, P.K. Bearing signature analysis as a medium for fault detection: A review. *J. Tribol.-Trans. ASME* **2008**, *130*, 1–6.

10. Li, B.; Chow, M.Y.; Tipsuwan, Y.; Hung, J.C. Neural network based motor rolling bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2000**, *47*, 1060–1069.

11. Al-Raheem, K.F.; Abdul-Karem, W. Rolling bearing fault diagnostics using artificial neural networks based on Laplace wavelet analysis. *Int. J. Eng. Sci. Technol.* **2010**, *2*, 278–290.

12. Samanta, B.; Al-Balushi, K.R. Artificial neural network based fault diagnostics of rolling element bearing using time domain features. *Mech. Syst. Signal Process.* **2003**, *17*, 317–328.

13. Gryllias, K.C.; Antoniadis, I.A. A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. *Eng. Appl. Artif. Intell.* **2012**, *25*, 326–344.

14. Konar, P.; Chattopadhyay, P. Bearing fault detection of induction motor using wavelet and support vector machines (SVMs). *Appl. Soft Comput.* **2011**, *11*, 4203–4211.

15. Zhang, X.Y.; Zhou, J.Z. Multi-fault diagnosis for rolling element bearing based on ensemble empirical mode decomposition and optimized support vector machines. *Mech. Syst. Signal Process.* **2013**, *41*, 127–140.

16. Anagnostopoulos, G.C.; Georgiopoulos, M. Ellipsoid ART and ARTMAP for incremental unsupervised and supervised learning. In Proceedings of Conference on Applications and Science of Computational Intelligence IV, Orlando, FL, USA, 17–18 April 2001; pp. 293–304.

17. Peralta, R.; Anagnostopoulos, G.C.; Gomez-Sanchez, E.; Richie, S. On the design of an ellipsoid ARTMAP classifier within the Fuzzy Adaptive System ART framework. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Montreal, QC, Canada, 31 July–4 August 2005; pp. 469–474.

18. Anagnostopoulos, G.C.; Georgiopoulos, M.; Verzi, S.J.; Heileman, G.L. Boosted ellipsoid ARTMAP. In proceedings of Conference on Applications and Science of Computational Intelligence V, Orlando, FL, USA, 2–3 April 2002; pp. 74–85.

19. Carpenter, G.A.; Martens, S.; Ogas, O.J. Self-organizing information fusion and hierarchical knowledge discovery: A new framework using ARTMAP neural networks. *Neural Netw.* **2005**, *18*, 287–295.

20. Skubalska-Rafajłowicz, E. Random projection RBF nets for multidimensional density estimation. *Int. J. Appl. Math. Comput. Sci.* **2008**, *18*, 455–464.

21. Xu, R.; Anagnostopoulos, G.C.; Wunsch, D.C. Tissue classification through analysis of gene expression data using a new family of ART architectures. In Proceeding of the 2002 International Joint Conference on Neural Networks, Honolulu, HI, USA, 12–17 May 2002; pp. 300–304.

22. Anagnostopoulos, G.C.; Georgiopoulos, M.; Verzi, S.J.; Heileman, G.L. Reducing generalization error and category proliferation in ellipsoid ARTMAP via tunable misclassification error tolerance: Boosted ellipsoid ARTMAP. In Proceedings of the 2002 International Joint Conference on Neural Networks, Honolulu, HI, USA, 12–17 May 2002; pp. 2650–2655.

23. Anagnostopoulos, G.C.; Georgiopoulos, M. Ellipsoid ART and ARTMAP for incremental clustering and classification. In proceedings of International Joint Conference on Neural Networks (IJCNN 01), Washington, DC, USA, 15–19 July 2001; pp. 1221–1226.

24. Das, S.; Suganthan, P.N. Differential evolution: A survey of the state-of-the-art. *IEEE Trans. Evolut. Comput.* **2011**, *15*, 4–31.

25. Stron, R. System design by constraint adaptation and Differential Evolution. *IEEE Trans. Evolut. Comput.* **1999**, *3*, 22–34.

26. Paterlini, S.; Krink, T. Differential evolution and particle swarm optimization in partitional clustering. *Comput. Stat. Data Anal.* **2006**, *50*, 1220–1247.

27. Luukka, P. Similarity classifier using similarities based on modified probabilistic equivalence relations. *Knowl.-Based Syst.* **2009**, *22*, 57–62.

28. Storn, R.; Price, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **1997**, *11*, 341–359.

29. Locatelli, M.; Maischberger, M.; Schoen, F. Differential evolution methods based on local searches. *Comput. Oper. Res.* **2014**, *43*, 169–180.

30. Jia, D.L.; Zheng, G.X.; Khan, M.K. An effective memetic differential evolution algorithm based on chaotic local search. *Inf. Sci.* **2011**, *181*, 3175–3187.

31. Al-Obeidat, F.; Belacel, N.; Carretero, J.A.; Mahanti, P. Differential evolution for learning the classification method PROAFTN. *Knowl.-Based Syst.* **2010**, *23*, 418–426.

32. He, W.; Jiang, Z.N.; Feng, K. Bearing fault detection based on optimal wavelet filter and space code shrinkage. *Measurement* **2009**, *42*, 1092–1102.

33. Neri, F.; Tirronen, V. Recent advances in differential evolution: A survey and experimental analysis. *Artif. Intell. Rev.* **2010**, *33*, 61–106.

34. Zarei, J.; Tajeddini, M.A.; Karimi, H.R. Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics* **2014**, *24*, 151–157.

35. Kiral, Z.; Karagulle, H. Vibration analysis of rolling element bearings with various defects under the action of an unbalanced force. *Mech. Syst. Signal Process.* **2006**, *20*, 1967–1991.

36. Liu, J.; Shao, Y.M.; Lim, T.C. Vibration analysis of ball bearing with a localized defect applying piecewise response function. *Mech. Mach. Theory* **2012**, *56*, 156–169.

37. Wu, T.; Yan, G.Z.; Yang, B.H.; Song, H. EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement* **2008**, *41*, 618–625.

38. Kandaswamy, K.K.; Pugalenthi, G.; Kalies, K.U.; Hartmann, E.; Martinetz, T. EcmPred: Prediction of extracellular matrix proteins based on random forest with maximum relevance minimum redundancy feature selection. *J. Theor. Biol.* **2013**, *317*, 377–383.

39. Li, B.Q.; Hu, L.L.; Niu, S.; Cai, Y.D.; Chou, K.C. Predict and analyze S-nitrosylation modification sites with the mRMR and IFS approaches. *J. Proteomics* **2012**, *75*, 1654–1665.

40. Stone, M. Cross-validatory choice and assessment of statistical prediction. *J. R. Stat. Soc. B Stat. Methodol.* **1974**, *36*, 111–147.