Mechanical Systems and Signal Processing 106 (2018) 526–536

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Intelligent fault recognition strategy based on adaptive optimized multiple centers

Bo Zheng [a,b], Yan-Feng Li [c], Hong-Zhong Huang [c,*]

[a] School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, PR China
[b] Civil Aviation Flight University of China, Guanghan, Sichuan 618307, China
[c] Center for System Reliability and Safety, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

ARTICLE INFO

Article history:
Received 29 July 2017
Received in revised form 5 November 2017
Accepted 14 December 2017

Keywords:
Multi-objective optimization
Priority levels
Fault recognition
Multiple centers
PSO variant

ABSTRACT

For the recognition principle based optimized single center, one important issue is that the data with nonlinear separatrix cannot be recognized accurately. In order to solve this problem, a novel recognition strategy based on adaptive optimized multiple centers is proposed in this paper. This strategy recognizes the data sets with nonlinear separatrix by the multiple centers. Meanwhile, the priority levels are introduced into the multi-objective optimization, including recognition accuracy, the quantity of optimized centers, and distance relationship. According to the characteristics of various data, the priority levels are adjusted to ensure the quantity of optimized centers adaptively and to keep the original accuracy. The proposed method is compared with other methods, including support vector machine (SVM), neural network, and Bayesian classifier. The results demonstrate that the proposed strategy has the same or even better recognition ability on different distribution characteristics of data.

1. Introduction

Fault recognition is a vital link of fault diagnosis, which is defined to determine the true states of the unknown fault patterns accurately. Actually, it has been widely and successfully applied to recognize running states of many complicated equipments, such as aeroengine, rotating machinery, and power electronic system [1–5]. It obviously helps to improve the efficiency of trouble-shoot, shorten the maintenance period, reduce the maintenance costs and ensure production safety. On the other hand, with the rapid development of industrial equipment, traditional analytical models can hardly describe real operations of complicated, huge and integrated equipment, and thus, the intelligence methods, such as including artificial neural network, statistical pattern recognition, swarm intelligence, and kernel-based algorithms, are paid more attentions, and have promoted the development of fault diagnosis technology in a more practical direction [6–8].

The advantage of intelligent fault recognition methods always focuses on dealing with the monitoring data of the running states without caring the physical or chemical running processes of equipment. Therefore, the methods can avoid the complicated mathematical description for the running process of the diagnosis object [9,10]. Meanwhile, with the development of intelligent computation, intelligent algorithms imitating natural phenomena of biological or physical systems can fully reflect the intelligent information processing mechanism contained in various systems, such as swarm intelligent algorithms simulating foraging behaviors of social animals [11,12], evolutionary algorithms deriving from the principle of the evolution-

* Corresponding author.
E-mail address: hzhuang@uestc.edu.cn (H.-Z. Huang).

https://doi.org/10.1016/j.ymssp.2017.12.026

ary laws in the biological world [13,14], and artificial intelligent algorithms based on behavior patterns, logic thinking and brain characteristics of human [15,16]. Undoubtedly, they have been widely used to solve various real-world problems, especially in the field of fault diagnosis.

The particle swarm optimization (PSO) algorithm is a classic of swarm intelligent algorithms, due to concise mathematical expression, good self-organization and adaptive performance, explicit individual interaction relationship. The PSO has developed rapidly and different well-known PSO variants were proposed to overcome the premature convergence problem caused by trapping into local suboptimal areas [17–19]. Furthermore, fitness functions, as a vital part of algorithm, are designed for specific applications, such as optimization design, feature extracting, path planning, and data clustering [20–23]. Although many studies on PSO always focus on its optimization applications, some studies try to apply PSO into pattern recognition, after the clustering algorithm based on PSO was proposed firstly in 2002 [24], the applications of PSO in pattern recognition and data mining increase rapidly [25–27]. In recent years, the applications of PSO have been extended to the field of prognostics and health management (PHM) [28–30]. Via combining with other classifiers, PSO variants improved the diagnosis and prognostics performances of classical classifiers, such as SVM [10,31–33], neural networks [7,34,35]. Zheng et al. [29,30] used the PSO variants to optimize a single center for each class, and the optimized single centers can meet the requirement of shorter intra-class distance, longer inter-class distance and maximum classification accuracy of training samples, so unknown samples can be recognized by comparing their distances with the optimized single centers. Virtually, the optimization of single center meeting three objectives is transformed to a single objective optimization problem. The significance of the studies is that the PSO becomes an independent classifier, not just an optimizer.

As a classifier, the experimental results [29,30] demonstrate that the performance of PSO variants is superior to other classical algorithms when applying to some data sets. Of course, no free lunch (NFL) theorem indicates that any pattern recognition algorithm cannot hold the superiority in its blood, it is impossible to be effective for all problems [36]. Consequently, the recognition principle based on distance from optimized single center to unknown samples has some obvious defects for some data sets. In other words, some data sets cannot be recognized accurately depending on only an optimized single center. Therefore, based on previous studies, this research will focus on the improvement of the principle based on optimized single center, and make the improved recognition method cope with more data sets.

In this paper, to solve the problems of the recognition principle based on the optimized single center, a novel recognition strategy is proposed. Moreover, using the optimization capability of PSO, a multi-objective optimization problem is constructed to meet more specific targets. Eventually, the accuracy is a prerequisite for fault recognition, so it must be met firstly as a primary objective, and on the basis of meeting first objective, other objectives, such as the quantity of optimized centers, the relationship of intra-class distances and inter-class distances, will be reconsidered and get new optimized results. Therefore, the strategy is based on adaptive optimized multiple centers.

On the other hand, the multi-team competitive optimization (MTCO) algorithm based on the traditional PSO has been demonstrated that it has globally stable and optimal performance [30], which is also a PSO variant. This algorithm is inspired by competitive behaviors of multiple teams. It is a three-level organization structure. And aim to searching more potential optimal areas, by imitating human thinking, the MTCO algorithm is conducive to get rid of the premature convergence effectively, and overcome the influence of randomness on the optimal decision solution. Thus the global optimal solution can be obtained with a higher probability. In this paper, it is noted that the MTCO algorithm is introduced as optimizer because of its better optimization and recognition performance. The detail description of MTCO algorithm can be found in Ref. [30].

The rest of the paper is organized as follows. In Section 2, the defects of the recognition principle based on optimized single center are analyzed. In Section 3, how to improve the appropriate optimized centers is discussed in detail. Section 4 compares the performances of the proposed method and some commonly used recognition methods, and proves the merits of multiple centers. After that, the proposed method is applied to the fault recognition to verify the effectiveness of multiple centers. Finally, in the last section, conclusions are drawn.

2. The defects of recognition principle based optimized single center

Zheng et al. [29,30] analyzed the performance of the recognition principle based on optimized single center, which has been demonstrated in Fig. 1. Obviously, the recognition principle can be classified into the method based on distance. In fact, k-nearest neighbors (k-NN) algorithm, learning vector quantization (LVQ) network, SVM, and so on, they are based on distance. The recent studies have indicated the defects of these algorithms, for examples, the k-NN is very sensitive to the parameter k, the different values of parameter k will severely affect the application of k-NN; and the defect of LVQ network has been discussed detailedly in Ref. [34] due to its non-identifiability for unknown samples and non-uniqueness for classification results; although the kernel function is introduced into SVM so that it can cope with the nonlinear distribution data, the common SVM can only recognize the two-class data [37]. Undoubtedly, the recognition principle based on optimized single center also suffers from defect.

As shown in Fig. 1, it depends on the distance from an unknown sample to the optimized centers, *i.e.*, if the distance from an unknown sample to the optimized center 1 is the shortest, this unknown sample should be classified into Class 1. Meanwhile, the methods are also verified by some classic data sets, and applied to the real-world fault recognition. The results show that the performance of the proposed methods is robust; their recognition accuracies are higher than other popular methods, such as SVM, and LVQ network [29,30].

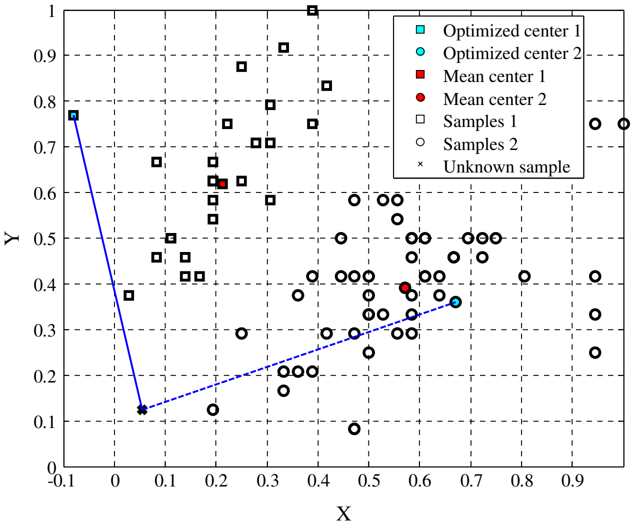
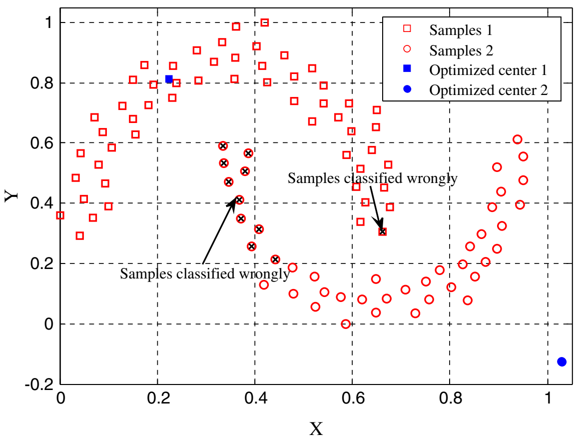
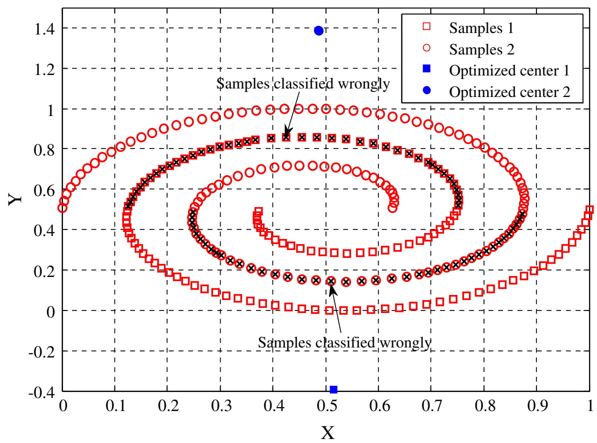


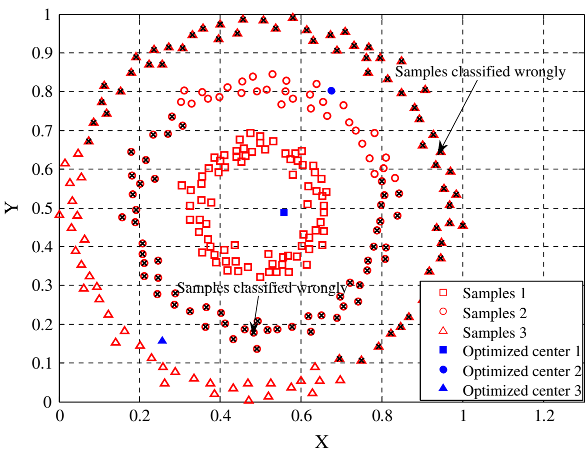
Fig. 1. The recognition principle based on optimized single center.



(a) Two-moons demo figure



(b) Spiral demo figure



(c) Three circles demo figure

Fig. 2. The recognition demo figures based on single center.

As mentioned above, due to only using the distances, for some data sets, it is difficult to find appropriate single center to recognize different patterns correctly. Considering visualizations, we take the 2-dimenional data as demo examples. Fig. 2 shows the 2-dimenional distribution of three data sets. Although the data in different classes have relatively obvious space interval, as shown in Fig. 2, the optimized single centers can hardly cope with the data sets similar to the three data sets. Basically, the data sets with obvious nonlinear separatrix cannot be recognized by only single optimized center. Essentially, as described in NFL theorem, no algorithms can be effective for all data sets; the recognition principles have determined the characteristics of algorithms. And the defects of the recognition principle based on optimized single centers are that it cannot recognize the data sets with nonlinear separatrix accurately. Undoubtedly, this situation limits the applications.

Accordingly, it is necessary to improve the recognition principle based on optimized single centers. Taking advantages of the PSO variant, the class centers can be optimized to recognize more data sets accurately.

3. Improvement of optimized centers

With the help of the optimization ability of PSO, the fitness functions for recognition are designed so that the single center can be optimized. A recognition method based PSO in Ref. [38] takes the shorter intra-class distance from training samples to their own class center as its fitness function. Based on the work in Ref. [38], the longer inter-class distance was also added into the fitness function in Ref. [39]. Furthermore, Zheng et al. [29,30] also considered the influence of the recognition accuracy of training samples on recognizing unknown samples, and given similar fitness function expressions based on the shorter intra-class distance, longer inter-class distance and maximum classification accuracy of training samples, respectively. And the calculation results indicated the fitness function, considering 3 factors, can improve the recognition accuracy more effectively. In fact, via transforming these three optimization objectives into one synthetical objective, the recognition methods based on PSO mentioned above were used to solve the single objective optimization problems.

3.1. The design of new fitness function

In this study, in order to overcome the defects discussed in Section 2, on the basis of the recognition principle based single centers, the new fitness function and new optimization objectives will be redesigned. The basic principle still depends on the distance from an unknown sample to the optimized centers, but in the new recognition strategy, the optimized centers for each category are not just single, and the quantity can be adjusted adaptively. Therefore, the recognition principle is based on adaptive optimized multi centers. To achieve the proposed effect, there are three objectives to be optimized, that is to say, this is a multi-objective optimization problem. Three objectives are written as follows:

$$\begin{cases} \max \mathrm{fit}_1(\boldsymbol{p}) = \frac{n_{\mathrm{trc}}}{n_{\mathrm{ts}}} \\ \min \mathrm{fit}_2(\boldsymbol{p}) = n_{\mathrm{toc}}, n_{\mathrm{toc}} = \sum_{l=1}^{n_c} n_{\mathrm{oc}}(l) \\ \max \mathrm{fit}_3(\boldsymbol{p}) = \frac{\mathrm{distances}_{\text{inter-centers}} + \mathrm{distances}_{\text{intra-centers}}}{\mathrm{distances}_{\text{samples-centers}}} \end{cases}, \tag{1}$$

where $\boldsymbol{p}$ denotes a particle, the position of $\boldsymbol{p}$ is feasible solution; $\mathrm{fit}_1(\cdot)$ represents the recognition accuracy of training samples, and n_{trc} is the number of training samples recognized correctly, n_{ts} is the total number of training samples; n_{toc} represents the total number of optimized centers, $n_{\mathrm{oc}}(l)$ indicates the number of optimized centers for the l-th class, and $l = 1, \ldots, n_c$, n_c indicates the total category number of data set; $\mathrm{fit}_3(\cdot)$ shows the distance relationship of optimized centers and training samples, inter-centers distances is defined as the distances from the optimized centers belonging to different classes, intra-centers distances is defined as the distances from the optimized centers belonging to the same class, and samples-centers distances is defined as the distances from training samples to their own optimized centers. Obviously, three distances are defined as follows:

$$\mathrm{distances}_{\text{inter-centers}} = \sum_{l=1}^{n_c-1} \sum_{k=l+1}^{n_c} \left(\left\{ \left\| \mathrm{mean}(\mathbf{c}_{li}) - \mathrm{mean}(\mathbf{c}_{kj}) \right\|_2 | i = 1, 2, \cdots, n_{\mathrm{oc}}(l), \quad j = 1, 2, \cdots, n_{\mathrm{oc}}(k) \right\} \right); \tag{2}$$

$$\mathrm{distances}_{\text{intra-centers}} = \sum_{l=1}^{n_c} \left(\min \left\{ \left\| \mathbf{c}_{li} - \mathbf{c}_{lj} \right\|_2 | i = 1, 2, \cdots, n_{\mathrm{oc}}(l) - 1, \quad j = i+1, \cdots, n_{\mathrm{oc}}(l), \right\} \right); \tag{3}$$

$$\mathrm{distances}_{\text{samples-centers}} = \sum_{i=1}^{n_{\mathrm{ts}}} \left(\min \left[\left\| \mathbf{x}_i - \mathbf{c}_{lj} \right\|_2 | j = 1, 2, \cdots, n_{\mathrm{oc}}(l), \quad cl(\mathbf{x}_i) = l \right] \right); \tag{4}$$

where $\boldsymbol{c}_{lj}$ denotes the j-th optimized center of the l-th class; $\boldsymbol{x}_i$ denotes the i-th training sample; and $\mathrm{cl}(\boldsymbol{x}_i)$ is used to show real class label of $\boldsymbol{x}_i$. And n_{trc} can be calculated by Eq. (5):

$$n_{\mathrm{trc}} = \sum_{i=1}^{n_{\mathrm{ts}}} \begin{cases} n_{\mathrm{trc}} + 1, & \left(cl(\mathbf{x}_i) = l \middle| \min\left[\left\| \mathbf{x}_i - \mathbf{c}_{lj} \right\|_2, l = 1, \cdots, n_{\mathrm{c}}, \quad j = 1, 2, \cdots, n_{\mathrm{oc}}(l) \right] \right) \\ n_{\mathrm{trc}} + 0, & \text{otherwise} \end{cases}. \tag{5}$$

Accordingly, this is a three-objective optimization problem, which meets the maximum recognition accuracy, minimum quantity of optimized centers, and maximum distance relationship. For $\mathrm{fit}_3(\cdot)$, it also comprehensively considers three-type distances, the inter-centers and intra-centers distances should be longer so that the optimized centers can be scattered in the space as far as possible, and the samples-centers distances should be shorter so that the samples recognized by the same center can be clustered as far as possible. As a result, $\mathrm{fit}_3(\cdot)$ can not only keep the distinction of all optimized centers, but also maintain the training samples identifiability, which can provide an accurate and clear guidance for recognition.

3.2. Multi-objective optimization based on priority levels

On the other hand, how to process the three-objective optimization problems will be taken into account. For solving the multi-objective optimization problems, the prevalent solving approach is always based on non-inferior solution, after solving the non-inferior solutions, the optimal solution can be determined by coordinating and selecting the non-inferior solutions [40]. In this study, aiming to the characteristics of PSO iterative optimization, the conception of priority levels is proposed, the qualitative description of priority levels includes "higher" and "lower". And thus, according to the actual demand of problem solving, each objective can be assigned a priority level; the objective with higher priority will be satisfied firstly, and then other objectives will be satisfied successively depending on their priority levels. For convenience, symbol "$\succ$" represents "be superior to", and symbol "$\nprec$" represents "be not worse than".

Assume m objectives need to be optimized, $F(\boldsymbol{p}) = [\mathrm{fit}_1(\boldsymbol{p}), \mathrm{fit}_2(\boldsymbol{p}), \ldots, \mathrm{fit}_m(\boldsymbol{p})]$, $\mathrm{P}(\cdot)$ is used to express the priority of a single objective. If $\mathrm{P}(\mathrm{fit}_1(\boldsymbol{p})) \succ \mathrm{P}(\mathrm{fit}_2(\boldsymbol{p})) \succ \ldots \succ \mathrm{P}(\mathrm{fit}_m(\boldsymbol{p}))$, the individual extremum $\boldsymbol{p}_{\mathrm{id}}$ and global extremum $\boldsymbol{p}_{\mathrm{gd}}$ can be updated by the following m conditions:

$$\begin{cases} 1 \cdot \begin{cases} \mathrm{fit}_1(\boldsymbol{p}) \succ \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{id}}) \\ \mathrm{fit}_1(\boldsymbol{p}) \succ \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{gd}}) \end{cases} \\ \vdots \\ m \cdot \begin{cases} \mathrm{fit}_1(\boldsymbol{p}) \nprec \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{id}}) \text{ and} \cdots \text{ and } \mathrm{fit}_{m-1}(\boldsymbol{p}) \nprec \mathrm{fit}_{m-1}(\boldsymbol{p}_{\mathrm{id}}) \text{ and } \mathrm{fit}_m(\boldsymbol{p}) \succ \mathrm{fit}_m(\boldsymbol{p}_{\mathrm{id}}) \\ \mathrm{fit}_1(\boldsymbol{p}) \nprec \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{gd}}) \text{ and} \cdots \text{ and } \mathrm{fit}_{m-1}(\boldsymbol{p}) \nprec \mathrm{fit}_{m-1}(\boldsymbol{p}_{\mathrm{gd}}) \text{ and } \mathrm{fit}_m(\boldsymbol{p}) \succ \mathrm{fit}_m(\boldsymbol{p}_{\mathrm{gd}}) \end{cases} \end{cases}. \tag{6}$$

If any of the conditions is satisfied, let $\boldsymbol{p}_{\mathrm{id}} = \boldsymbol{p}$, $\boldsymbol{p}_{\mathrm{gd}} = \boldsymbol{p}$, thus, the extremum can be realized. And for the adaptive optimized multiple centers, the priority levels are assigned like this: $\mathrm{P}(\mathrm{fit}_1(\boldsymbol{p})) \succ \mathrm{P}(\mathrm{fit}_2(\boldsymbol{p})) \succ \mathrm{P}(\mathrm{fit}_3(\boldsymbol{p}))$. Naturally, $\mathrm{fit}_1(\cdot)$ is the first objective, $\mathrm{fit}_2(\cdot)$ is the second objective, and $\mathrm{fit}_3(\cdot)$ is the third objective, based on these priority levels, $\boldsymbol{p}_{\mathrm{id}}$ and $\boldsymbol{p}_{\mathrm{gd}}$ will be updated according to the follow three conditions:

$$1 \begin{cases} \mathrm{fit}_1(\boldsymbol{p}) > \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{id}}) \\ \mathrm{fit}_1(\boldsymbol{p}) > \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{gd}}) \end{cases}; \tag{7}$$

$$2 \begin{cases} \mathrm{fit}_1(\boldsymbol{p}) \geqslant \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{id}}) \text{ and } \mathrm{fit}_2(\boldsymbol{p}) < \mathrm{fit}_2(\boldsymbol{p}_{\mathrm{id}}) \\ \mathrm{fit}_1(\boldsymbol{p}) \geqslant \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{gd}}) \text{ and } \mathrm{fit}_2(\boldsymbol{p}) < \mathrm{fit}_2(\boldsymbol{p}_{\mathrm{gd}}) \end{cases}; \tag{8}$$

$$3 \begin{cases} \mathrm{fit}_1(\boldsymbol{p}) \geqslant \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{id}}) \text{ and } \mathrm{fit}_2(\boldsymbol{p}) \leqslant \mathrm{fit}_2(\boldsymbol{p}_{\mathrm{id}}) \text{ and } \mathrm{fit}_3(\boldsymbol{p}) > \mathrm{fit}_3(\boldsymbol{p}_{\mathrm{id}}) \\ \mathrm{fit}_1(\boldsymbol{p}) \geqslant \mathrm{fit}_1(\boldsymbol{p}_{\mathrm{gd}}) \text{ and } \mathrm{fit}_2(\boldsymbol{p}) \leqslant \mathrm{fit}_2(\boldsymbol{p}_{\mathrm{gd}}) \text{ and } \mathrm{fit}_3(\boldsymbol{p}) > \mathrm{fit}_3(\boldsymbol{p}_{\mathrm{gd}}) \end{cases}; \tag{9}$$

Consequently, on the basis of priority levels, these optimization objectives can be met in proper order, specifically, this strategy ensures the maximum recognition accuracy is met firstly, and then the minimum quantity of optimized centers is just decided adaptively on the premise that the accuracy is not reduced; finally, the maximum distance relationship is also met only under the condition that the first two objectives don't become poor. Therefore, this recognition principle is based on adaptive optimized multiple centers.

For different characteristic data sets, due to the class centers being determined adaptively, if all $n_{\mathrm{oc}}(l) = 1$ ($l = 1, \ldots, n_{\mathrm{c}}$), multiple centers will be converted into the single center, which means that adaptive centers can continue to maintain the advantage of the recognition principle based on the optimized single center. On the other hand, if any $n_{\mathrm{oc}}(l) \neq 1$ ($l = 1, \ldots, n_{\mathrm{c}}$), which means some classes possess multiple centers, the advantage of multi-class centers will help to recognize the data sets with nonlinear separatrix accurately. Thus, the proposed method can expand the application so that more data types can be recognized.

With the help of MTCO algorithm, the procedure of the recognition strategy based on adaptive optimized multiple centers can be described as follows:

Step 1: The training and test samples should be normalized to be on interval [0, 1], which helps to eliminate influence of dimension, and reduce the search area and improve the algorithm efficiency.
Step 2: Initialize each particle's initial velocity and position in the whole solution space randomly. Each position is a potential optimal solution.
Step 3: The particle velocity and position are updated according to the MTCO algorithm, and the fitness value of the particles is calculated.
Step 4: The individual extremum $\boldsymbol{p}_{\mathrm{id}}$ and global extremum $\boldsymbol{p}_{\mathrm{gd}}$ can be updated according to fitness value change of population and the priority levels of objectives.
Step 5: Return Step 3 and repeat the iteration until the terminal conditions are satisfied.

4. Recognition experiment and performance comparison

4.1. Performance verification and comparison

In order to verify the recognition performance of the proposed method, several classic and typical data sets with obvious nonlinear separatrix are used; they are two moons, spiral, and three circles. In addition, some other data sets are also used; they are Wine, Indian liver patients, Wisconsin diagnostic breast cancer, and Seeds. So the recognition performance can be verified comprehensively. And the training and test samples are selected randomly; the initial class centers of the data sets are set as 15, which are shown in Table 3. In this study, some commonly used recognition methods, such as support vector machine (SVM) [28], back propagation (BP) network [41], learning vector quantization (LVQ) network [42], Bayes classifier based on expectation maximization (EM) and mixture normal distribution [43], and single center-based recognition method [30] are used to compare with the proposed method. The information of data sets are shown in Table 1.

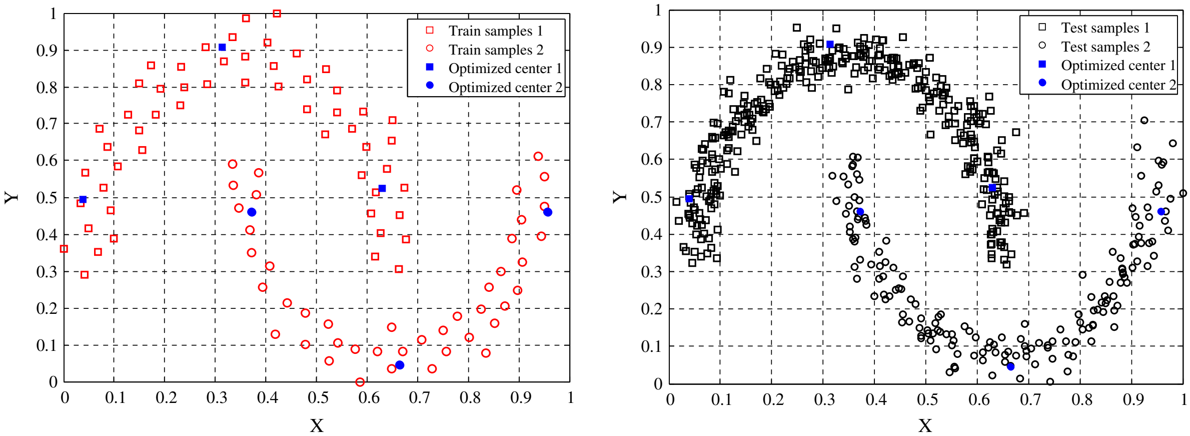
The settings of MTCO algorithm is like that the population size is 60, the total iteration number is 200, and other parameters are set as same as the settings in [30]. After calculating the three 2-dimensional data sets with nonlinear separatrix, Fig. 3 compares the optimization effects between single center and multiple centers, which demonstrates the effectiveness and practicality of recognition method based on optimized multi-class centers. Obviously, for the same three data sets, the proposed method can recognize the data sets with nonlinear separatrix accurately. Meanwhile, Fig. 4 indicates the multi-objective fitness values change of the three 2-dimensional data sets, and the change trends of fitness values reflect the optimization objectives are achieved, in which every objective is changed according to the designed ideas based on priority levels.

The rest of data sets are also calculated for performance comparison. Moreover, the randomness is the main factor affecting the performance of the MTCO algorithm, which may lead to obtain the uncertainty of results. Meanwhile, the influence of weight setting randomly on BP and LVQ always leads to the uncertainty of results. In addition, the kernel parameters of SVM optimized by the method proposed in [28] and the estimator of parameters based EM are still keeping uncertainly. In order to have a statistical soundness of the results and verify the ability of recognition, we did 20 experiments repeatedly under the same setting to get the statistical parameters of accuracies including Min, Mean, Max, and STD (Standard Deviation). Figs. 3 and 4 show one of 20 experimental results. For more details, please refer to the corresponding references cited above for setting the relative parameters and network structures of SVM, BP and LVQ network. Especially, the error goals for BP and LVQ network are set as 0.001, and the iteration numbers for them are set as 500 and 200, respectively. The number of mixture normal distribution is set as 3, the error goal is 10^{-6} and the iteration numbers of EM is set as 3000. All the statistical results are shown in Table 2.

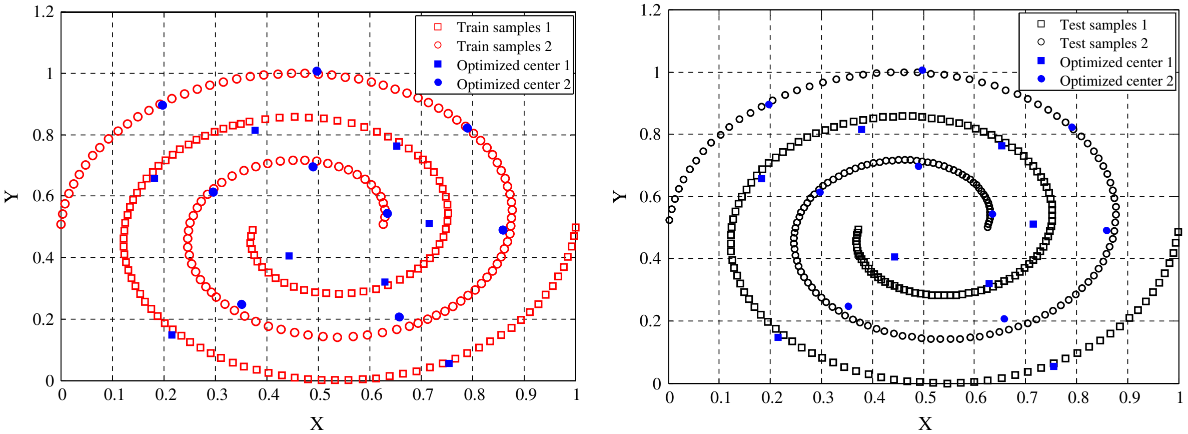
As shown in Table 2, some conclusions can be drawn by the recognition results of different data sets:

Table 1
The relative information about data sets.

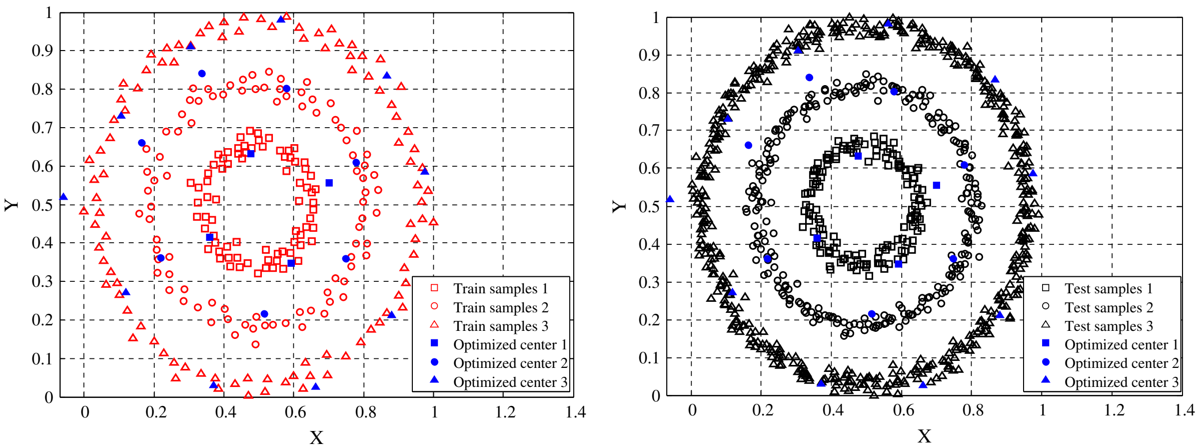
Name	Dimensions	Categories	Training samples	Test samples
Two moons	2	2	98 (55 + 43)	500 (333 + 167)
Spiral	2	2	244 (123 + 121)	314 (157 + 157)
Three circles	2	3	263 (77 + 88 + 98)	901 (150 + 251 + 500)
Wine	13	3	98 (34 + 38 + 26)	80 (25 + 33 + 22)
Indian liver patients (ILP)	10	2	213 (106 + 107)	120 (60 + 60)
Wisconsin diagnostic breast cancer (WDBC)	9	2	503 (324 + 179)	180 (120 + 60)
Seeds	7	3	120 (40 + 40 + 40)	80 (30 + 30 + 30)



(a) The optimized multi-class centers for two moons



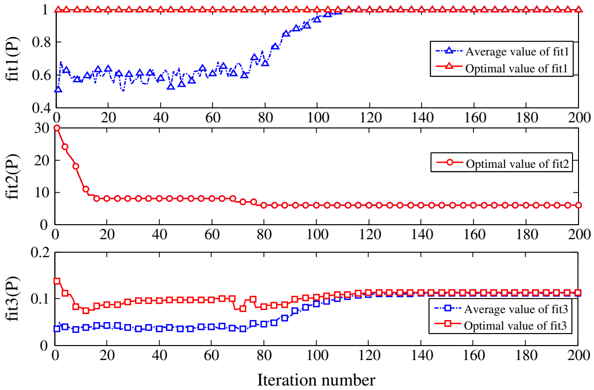
(b) The optimized multi-class centers for spiral



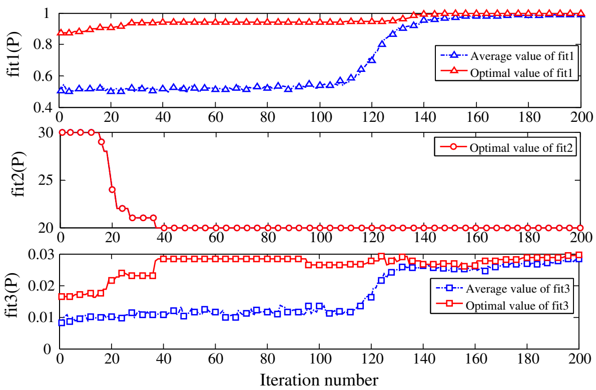
(c) The optimized multi-class centers for three circles

Fig. 3. The recognition demo figures based on multi-class centers.

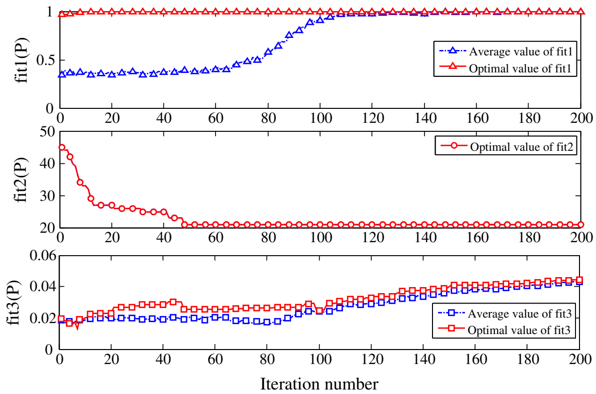
(1) For two moons, spiral, and three circles, the recognition accuracies of multi-class centers are obviously better than theses of single class center. Compared with other algorithms, the performance of multiple centers only slightly inferior to that of SVM, in most cases, its performance is better than other popular algorithms, such as BP, LVQ, Bayses classifier.



(a) Fitness values change for two moons



(b) Fitness values change for spiral



(c) Fitness values change for three circles

Fig. 4. Trends of fitness value change.

(2) For other data sets, the recognition performance of multi-class centers is basically equal to that of single class center, even better than it. Although the recognition principles of other popular algorithms are different, they can only work on some of the data sets, but the proposed principle is almost effective for all data sets.

Undoubtedly, the recognition principle based on multiple centers retains all benefits of the single center principle; what's more, it expands the application range and can deal with the data with nonlinear separatrix effectively. So the improvement strategy proposed in this paper is proven effective.

The main reason for the results shown in Table 2 is that the principles of different algorithms have their own focuses, which can only solve a specific type of problems. SVM relies on nonlinear hyper-plane mapped by kernel functions; BP network relies on the nonlinear mapping of transfer functions; LVQ network relies on the distance to the winner neurons; and Bayes takes the mixture normal distribution of training samples as prior probability distribution. For these data with obvious between-class space interval, they can be recognized easily by SVM and BP, so for the two moons, spiral, and three circles, they have good accuracy. However, the performance of neural network is easily affected by their random setting weights. Although mixture normal distribution can describe the distribution of unknown probability density function in theory, Bayes classifier cannot recognize the data with complicated distributions; it is still sensitive to the different probability distributions of data, which is more suitable for these data with normal distributions. Similar to the single center principle, LVQ is not inappropriate for the data with obvious nonlinear separatrix, on the other hand, these recognition principles based on the distances are effective for data with irregular space distributions; for the ILP, they have relatively good performances. To summarize, the results in Table 2 indicate that the proposed method has good recognition ability to different distribution characteristics of data, and randomness has made little or even no effect on the proposed method. Obviously, it can realize the accurate recognition to the commonly used fault pattern data.

On the other hand, due to determining the quantity of optimized centers adaptively, in order to verify the adaptive performance, Table 3 indicates the quantity statistics of optimized centers for 7 data sets. It demonstrates the adaptive capability of determining the optimized centers according to different data characteristics. Therefore, minimum quantity of optimized centers is just obtained adaptively on the premise that the accuracy is not reduced.

Table 2
Recognition accuracy comparison of multi-center method and other algorithms.

Data set	Algorithms	Min	Mean	Max	STD
Two moons	Multiple centers	1	1	1	0
	Single center	0.8360	0.8372	0.8480	0.0038
	BP	1	1	1	0
	LVQ	0.9940	0.9956	0.9980	0.0010
	SVM	1	1	1	0
	Bayes	0.8700	0.9545	0.9800	0.0343
Spiral	Multiple centers	0.9936	0.9994	1	0.0020
	Single center	0.6656	0.6656	0.6656	0
	BP	1	1	1	0
	LVQ	0.6401	0.6546	0.6783	0.0088
	SVM	1	1	1	0
	Bayes	0.6178	0.7065	0.7580	0.0303
Three circles	Multiple centers	0.9967	0.9990	1	0.0011
	Single center	0.3807	0.4744	0.6570	0.0983
	BP	0.9654	0.9894	1	0.0104
	LVQ	0.6426	0.6925	0.7758	0.0370
	SVM	1	1	1	0
	Bayes	0.7980	0.9095	0.9767	0.0471
Wine	Multiple centers	0.9625	0.9772	1	0.0091
	Single center	0.9625	0.9737	0.9750	0.0040
	BP	0.8625	0.9306	0.9875	0.0310
	LVQ	0.9000	0.9275	0.9500	0.0132
	SVM	0.9625	0.9748	0.9875	0.0031
	Bayes	0.9625	0.9735	0.9875	0.0071
ILP	Multiple centers	0.6833	0.7165	0.7583	0.0171
	Single center	0.6417	0.7017	0.7333	0.0266
	BP	0.2667	0.4165	0.5417	0.0621
	LVQ	0.5417	0.6700	0.7250	0.0451
	SVM	0.4250	0.4393	0.4583	0.0106
	Bayes	0.4333	0.6051	0.7000	0.0531
WDBC	Multiple centers	0.9611	0.9611	0.9611	0
	Single center	0.9611	0.9611	0.9611	0
	BP	0.8667	0.9214	0.9500	0.0157
	LVQ	0.9000	0.9181	0.9389	0.0114
	SVM	0.9611	0.9611	0.9611	0
	Bayes	0.9444	0.9464	0.9611	0.0045
Seeds	Multiple centers	0.9556	0.9683	0.9778	0.0065
	Single center	0.9556	0.9627	0.9778	0.0080
	BP	0.9333	0.9531	0.9667	0.0115
	LVQ	0.9444	0.9656	0.9778	0.0088
	SVM	0.9556	0.9556	0.9556	0
	Bayes	0.9333	0.9484	0.9667	0.0118

Table 3
The statistics of optimized centers.

Name	Categories	Initial centers	Statistics of optimized centers			
			Max	Mean	Min	STD
Two moons	2	$2 \times 15 = 30$	6	6	6	0
Spiral	2	$2 \times 15 = 30$	25	23	20	1.5
Three circles	3	$3 \times 15 = 45$	26	22.5	18	2.0286
Wine	3	$3 \times 15 = 45$	7	4.9778	3	1.0551
ILP	2	$2 \times 15 = 30$	2	2	2	0
WDBC	2	$2 \times 15 = 30$	2	2	2	0
Seeds	3	$3 \times 15 = 45$	3	3	3	0

4.2. A bleed air fault recognition of a certain aeroengine

We take the bleed air fault of a certain aeroengine as diagnostic object. The obvious change of status monitoring parameters, mainly including EGT (Exhaust Gas Temperature), FF (Fuel Flow), N2 (High pressure rotors speed), collected from the actual flights of aircraft indicate the health of aeroengine. Through the analysis of the monitored data and the confirmation

Table 4
Fault message of the aeroengine.

	M	PA (fleet)	TAT (°C)	TLA (°)	FF (lb/h)	N1 (%)	N2 (%)	EGT (°C)	T2 (°C)	P2 (psia)	D
Train samples	0.746	34,082	−25.1	64.9	1200	85.82	91.69	712	4	7.8	1
	0.726	26,578	−6.1	67.7	1568	86.62	94.24	736	23.7	10.5	1
						⋮					
	0.744	34,094	−11.8	68.1	1300	88.71	94.7	787	19.7	7.8	2
	0.739	31,078	−3.2	75.2	1448	89.13	95.73	795	28.6	8.8	2
Test samples	0.746	27,584	1.4	69.6	1640	88.12	96.27	803	31.5	10.3	1
						⋮					
	0.725	26,576	−1.4	67.9	1632	87.44	95.45	774	29	10.5	2

Table 5
Diagnosis comparison for bleed air fault.

Algorithm	Min	Mean	Max	STD
Multiple centers	0.8235	0.8353	0.8824	0.0241
Single center	0.8235	0.8265	0.8824	0.0300
BP	0.7059	0.7912	0.8235	0.0356
LVQ	0.6471	0.6941	0.7647	0.0452
SVM	0.7647	0.8235	0.8824	0.0427
Bayes	0.5882	0.6891	0.7647	0.0422

from the ground inspection, the status change of aeroengine is caused by bleed air fault. Thus, the collection and management of this air bleed fault is critical for the future fault diagnosis.

According to the flight environment, performance parameter, a fault message table aiming to this fault is organized in Table 4. There are 10 attributes, they are Mach (M), pressure altitude (PA), total air temperature (TAT), throttle lever angle (TLA), fuel flow (FF), exhaust gas temperature (EGT), N1 (Low pressure rotors speed), N2 (High pressure rotors speed), compressor delivery temperature (T2), compressor delivery pressure (P2), respectively. And there are only two statuses including normal and bleed air fault, respectively, where ‘1’ denotes normal and ‘2’ denotes bleed air fault. 19 samples are selected as train samples randomly, 11 samples are normal, and 8 samples are bleed air fault. The rest of 17 samples are test samples, in which 13 samples are normal, and 4 samples are bleed air fault.

Actually, small samples can test the performance of the algorithms better due to limited information support [44]. We still do 20 experiments to compare the performances of the algorithms used in Section 4.1; all the settings are the same as Section 4.1. The results are shown in Table 5. The comparison results clearly indicate the proposed recognition strategy is still effective for the real-world fault pattern recognition problem, and the recognition results for the unknown fault pattern prove the better performance of the proposed strategy.

5. Conclusions

In this paper, in order to overcome the defect of the recognition principle based on optimized single center, a novel recognition strategy based on adaptive optimized multi-center is proposed. According to the above experimental results, main conclusions obtained in this study are summarized as follows:

(1) Taking advantage of multi-center, a novel strategy is proposed to cope with these data sets with nonlinear separatrix accurately. The quantity of the centers can be adjusted adaptively according to different data characteristics; and thus the proposed strategy not only maintains the original merits of the recognition principle based on single center, but also expands its scope of applications.
(2) The quantity of optimized centers is adaptively adjusted on the premise that the accuracy is not reduced. And the distance relationship is also met only under the condition that the first two objectives don’t become poor. So the three-objective optimization base on priority levels is designed to ensure better performances of the optimized centers.
(3) Although this paper doesn’t develop a new PSO variant, multi-team competitive the optimization algorithm is used as an optimization tool. The MTCO algorithm is proven to have excellent performance on solving single-objective optimization. In this study, it further solve the problem of the multi-objective optimization.

Acknowledgements

This research was supported by the Fundamental Research Funds for the Central Universities under contract numbers ZYGX2014Z010 and SKLMT-KFKT-201601.

References

[1] H.Z. Huang, P. Cui, W. Peng, H. Gao, H.K. Wang, Fatigue lifetime assessment of aircraft engine disc via multi-source information fusion, Int. J. Turbo Jet Engines 31 (2) (2014) 167–174.
[2] J. Yang, H.Z. Huang, R. Sun, H. Wan, Reliability analysis of aircraft servo-actuation systems using evidential networks, Int. J. Turbo Jet Engines 29 (2) (2012) 59–68.
[3] Y.F. Li, H.Z. Huang, S.P. Zhu, Y. Liu, N.C. Xiao, An application of fuzzy fault tree analysis to uncontained events of an areo-engine rotor, Int. J. Turbo Jet Engines 29 (4) (2012) 309–315.
[4] L. Gan, H.Z. Huang, S.P. Zhu, Y.F. Li, Y. Yang, Fatigue reliability analysis of turbine disk alloy using saddlepoint approximation, Int. J. Turbo Jet Engines 30 (3) (2013) 217–230.
[5] H.Z. Huang, J. Gong, M.J. Zuo, S.P. Zhu, Q. Liao, Fatigue life estimation of an aircraft engine under different load spectrums, Int. J. Turbo Jet Engines 29 (4) (2012) 259–267.
[6] Y.F. Li, H.Z. Huang, H. Zhang, N.C. Xiao, Y. Liu, Fuzzy sets method of reliability prediction and its application to a turbocharger of diesel engines, Adv. Mech. Eng. 2013 (4) (2013) 1–7.
[7] H.Z. Huang, Y. Liu, Y. Li, L. Xue, Z. Wang, New evaluation methods to concept selection using computational intelligence technique, J. Mech. Sci. Technol. 27 (3) (2013) 733–746.
[8] Y.F. Li, H.Z. Huang, Y. Liu, N. Xiao, H. Li, A new fault tree analysis method: fuzzy dynamic fault tree analysis, Eksploatacja i Niezawodnosc – Mainten. Reliab. 14 (3) (2012) 208–214.
[9] B. Zheng, H.Z. Huang, Y.F. Li, Aeroengine fault diagnosis method based on optimized supervised Kohonen network, J. Donghua Univ. 32 (6) (2015) 1029–1033.
[10] H.Z. Huang, H.K. Wang, Y.F. Li, L. Zhang, Z. Liu, Support vector machine based estimation of remaining useful life: current research status and future trends, J. Mech. Sci. Technol. 29 (1) (2015) 151–163.
[11] J. Kennedy, R.C. Eberhart, Particle swarm optimization, in: Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ, 1995, pp. 1941–1948.
[12] S. Binitha, S.S. Sathya, A survey of bio inspired optimization algorithm, Int. J. Soft Comput. Eng. 2 (2) (2012) 137–151.
[13] S. Droste, T. Jansen, G. Rudolph, H.P. Schwefel, K. Tinnefeld, I. Wegener, Theory of evolutionary algorithms and genetic programming, Nat. Comput. 29 (1) (2003) 107–144.
[14] T. Back, U. Hammel, H.P. Schwefel, Evolutionary computation: comments on the history and current state, IEEE Trans. Evolut. Comput. 1 (1) (2002) 3–17.
[15] S.J. Russell, P. Norvig, Artificial intelligence: a modern approach, Appl. Mech. Mater. 263 (5) (2010) 2829–2933.
[16] A. Bhardwaj, A. Tiwari, Breast cancer diagnosis using Genetically Optimized Neural Network model, Expert Syst. Appl. 42 (10) (2015) 4611–4620.
[17] R. Cheneg, Y.C. Jin, A social learning particle swarm optimization algorithm for scalable optimization, Inf. Sci. 291 (6) (2015) 43–60.
[18] M.R. Tanweer, S. Suresh, N. Sundararajan, Self regulating particle swarm optimization algorithm, Inf. Sci. 294 (10) (2015) 182–202.
[19] S. Janson, M. Middendorf, A hierarchical particle swarm optimizer and its adaptive variant, IEEE Trans. Cybern. 35 (6) (2006) 1272–1282.
[20] Y. Song, Z. Chen, Z. Yuan, New chaotic PSO-based neural network predictive control for nonlinear process, IEEE Trans. Neural Netw. 18 (2) (2007) 595–601.
[21] S. Suresh, P.B. Sujit, A.K. Rao, Particle swarm optimization approach for multi-objective composite box-beam design, Compos. Struct. 81 (4) (2008) 598–605.
[22] A.K. Husseinzadeh, M.K. Husseinzadeh, S. Karimiyan, A particle swarm optimizer for grouping problems, Inf. Sci. 252 (17) (2013) 81–95.
[23] L. Cagnina, M. Errecalde, D. Ingaramo, P. Rosso, An efficient particle swarm optimization approach to cluster short texts, Inf. Sci. 265 (5) (2014) 36–49.
[24] M.G. Omrang, A. Salman, A.P. Engelbrecht, Image classification using particle swarm optimization, in: Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning, 2002, pp. 370–374.
[25] C.W. Tsai, K.W. Huang, C.S. Yang, M.C. Chiang, A fast particle swarm optimization for clustering, Soft. Comput. 19 (2) (2015) 321–338.
[26] X. Zhang, L. Jiao, A. Paul, Y. Yuan, Z. Wei, Q. Song, Semisupervised particle swarm optimization for classification, Math. Probl. Eng. 2 (2014) 1–11.
[27] C.X. Wang, Novel model of particle swarm optimization for data mining based on improved ant colony algorithm, J. Chem. Pharma. Res. 6 (8) (2014) 190–197.
[28] B. Zheng, Investigation on aeroengine maintenance level decision on PSO-SVM, J. Propuls. Technol. 34 (5) (2013) 687–692.
[29] B. Zheng, F. Gao, Fault diagnosis method based on S-PSO classification algorithm, Acta Aeronautica et Astronautica Sinica 36 (11) (2015) 3640–3651.
[30] B. Zheng, H. Z. Huang, H.W. Xu, D.B. Meng, X.L. Zhang, Multi-team competitive optimization algorithm and its application in bearing fault diagnosis, in: Annual Reliability and Maintainability Symposium, 2016.
[31] B. Zheng, F. Gao, Research on the prediction of aeroengine wear based on the IPSO-SVR, Lubr. Eng. 39 (11) (2014) 81–87.
[32] S.F. Yuan, F.L. Chu, Fault diagnostics based on particle swarm optimization and support vector machines, Mech. Syst. Sign. Process. 21 (4) (2007) 1787–1798.
[33] F.F. Chen, B.P. Tang, T. Song, L. Li, Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle swarm optimization, Measurement 47 (1) (2014) 576–590.
[34] B. Zheng, Y.F. Li, H.Z. Huang, Aeroengine fault diagnosis method based on optimized supervised Kohonen network, J. Donghua Univ. 32 (6) (2015) 1029–1033.
[35] V. Fathi, G.A. Montazer, An improvement in RBF learning algorithm based on PSO for real time applications, Neurocomputing 111 (6) (2013) 169–176.
[36] C.J. D'Orsi, Computer-aided detection: there is no free lunch, Radiology 221 (3) (2001) 585–586.
[37] H.J. Cheng, L.C. Chen, Y.J. Zhang, J. Li, On research of algorithms about structuring multi-Classifier based on SVM, Comput. Technol. Dev. 18 (12) (2008) 109–112.
[38] S. Kiranyaz, T. Ince, A. Yildirim, Evolutionary artificial neural networks by multi-dimensional particle swarm optimization, Neural Netw. 22 (10) (2009) 1448–1462.
[39] S.J. Nanda, G. Panda, Automatic clustering algorithm based on multi-objective immunized PSO to classify actions of 3D human models, Eng. Appl. Artif. Intell. 26 (5–6) (2013) 1429–1441.
[40] N. Wang, W.J. Zhao, N. Wu, D. Wu, Multi-objective optimization: a method for selecting the optimal solution from Pareto non-inferior solutions, Expert Syst. Appl. 74 (2017) 96–104.
[41] J.B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, F. Fnaiech, Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration, Appl. Acoust. 89 (3) (2015) 16–27.
[42] B. Biswal, M. Biswal, S. Hasan, P.K. Dash, Nonstationary power signal time series data classification using LVQ classifier, Appl. Soft Comput. 18 (1) (2014) 158–166.
[43] M.M. Lange, N.A. Novikov, Bayes classifier based on tree-structured Gaussian mixtures, Pattern Recogn. Image Anal. 22 (1) (2012) 136–143.
[44] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Disc. 2 (2) (1998) 121–167.