Springer

KSME

Journal of Mechanical Science and Technology 33 (1) (2019) 129~138
www.springerlink.com/content/1738-494x(Print)/1976-3824(Online)
DOI 10.1007/s12206-018-1213-6

Sharing pattern feature selection using multiple improved genetic algorithms and its application in bearing fault diagnosis [†]

Xiaoying Guan[1,2] and Guo Chen[1,*]

[1]*College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[2]*School of Software, Guangdong Food and Drug Vocational College, Guangzhou 510000, China*

(Manuscript Received March 28, 2018; Revised July 24, 2018; Accepted August 14, 2018)

Abstract

In order to select the effective features or feature subsets and realize an intelligent diagnosis of aero engine rolling bearing faults, this paper presents a sharing pattern feature selection method using multiple improved genetic algorithms. Based on the simple genetic algorithm, a multiple-population improved genetic algorithm was proposed, which improves the speed and effect of algorithm and overcomes the shortcomings of local optima that simple genetic algorithm is easy to fall into. Because all populations regularly share and exchange their selecting features, the proposed algorithms can quickly dig up the current effective feature patterns, and then analyze and deal with the strong correlation between the feature patterns. This will not only give clear directions for the descendant evolution, but also help to achieve high accuracy feature selection, for, the features are highly distinctive. This multiple-population improved genetic algorithm was applied to rolling bearing fault feature selection and comparisons with other methods are carried out, which demonstrates the validity of sharing pattern feature selection method proposed.

Keywords: Feature selection; Feature pattern; Multiple-population; Genetic algorithm; Bearing; Fault diagnosis

---

1. Introduction

As the key component of aero-engine, aero-engine rolling bearing has a direct impact on the flight safety. In order to realize the rolling bearing condition monitoring, it is necessary to collect the environment data and operating data of rolling bearing. In the aspect of feature extraction, there are wildly used time-domain features such as peak value, root mean square, shape indicator, crest indicator, impulse indicator, clearance indicator, kurtosis and skewness [1-3]. As for frequency domain feature extraction, the widely used methods are amplitude spectrum analysis, power spectrum analysis, cepstrum analysis, high-order spectrum analysis and envelope spectrum analysis [4-8]. Feature extraction methods based on time-frequency analysis include short-time Fourier transform, Wigner-Ville distribution, wavelet and wavelet packet analysis, empirical mode decomposition, spectral kurtosis and singular value decomposition [9-12]. Because the single feature is susceptible as well as the sensitivity of different features of different faults are different, bearing condition assessment based on a single feature is not ideal enough, and it is hard to be applied in practical engineering. A good condition assessment method is supposed to be robustness to working condition and sensitive to all kinds of faults. Therefore, the multi-dimensional feature information of time domain, frequency domain and time-frequency domain should be fully utilized in the fault diagnosis of rolling bearing. However, not all of the features play a positive role in bearing state discrimination. We hope to select effective feature subset, which not only can effectively distinguish the bearing state as much as possible, but also make the difference between the features of this feature subset large. Therefore, it is necessary to take effective method to select the best feature subset from the multi-dimensional features and to achieve high accuracy of the rolling bearing fault diagnosis. This problem can be achieved by the feature selection in pattern recognition.

The task of feature selection is to find a set of the most effective features for classification, which requires that when the feature dimension is reduced to the same level, its classification performance is the best. This is a process to solve the problem and find optimized combinatorial solutions. In this respect, genetic algorithm (GA) has great potential, especially when the choice space is large (high feature dimensions) and lack of understanding of the relationship between features. Genetic algorithm is a randomized optimization algorithm with orientation, which is used to simulate the genetic evolution of biological evolution in nature. It has good global search

---

[*] Corresponding author. Tel.: +86 25 84891850, Fax.: +86 25 84891850
E-mail address: cgzyx@263.net

[†] Recommended by Associate Editor Gyuhae Park

ability and implicit parallelism; therefore, it is widely used in feature selection.

However, although genetic algorithm has strong global optimization potential, it is also easy to fall into local optima. This problem can be realized mainly through enhancing the genetic operators (selection operators, crossover operators and mutation operators) to improve the performance of the algorithm, with higher efficiency to find better features or subset [13-16]. For example, Yang et al. [14] proposed a criterion function for selecting the optimal feature subset and a search strategy called novel quantum genetic algorithm (NQGA). NQGA adopted a novel update approach of rotation angles of quantum gates, and immigration and catastrophe operations to enhance search capability and to avoid premature convergence. The proposed feature selection algorithm reduces greatly the dimensions of original feature set and heightens accurate recognition rate of radar emitter signals. Li et al. [15] proposed a multi-population agent genetic algorithm with double chain-like agent structures (close chain-like agent structure and cycle chain-like agent structure) for feature selection, with dynamic neighborhood competition selection operator, neighborhood orthogonal crossover operator and adaptive mutation operator to effectively keep and enhance the diversity of the subpopulation, which is good to search for global optima in complex and high dimensional search space.

Genetic algorithm can also be applied to the feature selection of the rolling bearing faults through state analysis and fault diagnosis. Chen et al. [17] presented a new method called “automated function generation of symptom parameters” using genetic algorithms (GA). The proposed method is effective to express the feature of the signal for discriminating between normal and abnormal states. And the mutation probability of GA is not based on a fixed probability but is obtained from the distinction index of the gene position. Kang et al. [18] presented a reliable fault diagnosis methodology for various single and multiple combined defects of low-speed rolling element bearings. The proposed reliable diagnosis approach employs a genetic algorithm (GA)-based discriminative feature analysis (GADFA) to select the most discriminative subset of all the extracted fault signatures, and the GA used in this paper focuses on the impact of population size. Ettefagh et al. [19] proposed a hybrid GA-K-mean clustering used in fault diagnosis of the scaled rotor bearing system experimentally, and simple genetic algorithm (SGA) is applied to overcome the drawback of K-means.

Rolling bearings, as a key component of the aero engine, directly affects flight safety, it is necessary to take an effective method to evaluate and select the state data of the rolling bearing, so as to guarantee an intelligent diagnosis. According to the above results, genetic algorithm is applied to the fault diagnosis of rolling bearings, but more is adopting a single population with genetic algorithm or simple genetic algorithm [20]. In contrast, when genetic algorithm is applied to multiple populations, the search speed and the quality of results will be improved due to the sharing and exchanging of stage results between populations, which will be further applied to guide the evolution of offspring. Thus, this paper proposes a multiple improved genetic algorithms with sharing the pattern (SPMIGA), which is applied to feature selection for roller bearing fault diagnosis. Firstly, the Improved Genetic Algorithm (IGA) is adopted based on SGA, which improves the mutation operator by using the variance operator. The IGA achieves the effect of speeding up the population convergence. And then, each of the multiple populations independently executes the IGA operations to search for the optimal feature subsets. After its evolution with short-iterations, each population gets the current best individual, and then uses it in searching for the effective feature pattern. Combined with the Pearson correlation coefficient and the weight values of the features obtained by ReliefF algorithm, SPMIGA can deal with feature selection with strong correlation. Then, each of the population reinitializes the individuals according to the feature pattern to generate a new population. This procedure continues until the satisfactory terminal feature pattern is reached.

2. ReliefF

ReliefF [21], an improved relief algorithm, is proposed by Kononenko in 1994. It is also one of the current means of feature selection under the filter method, which is mainly used to deal with many types of problems and regression problems. By constantly adjusting the values attributed to the features, the features that correlate to the category are given a higher weight.

ReliefF randomly selects an instance R_i from the training sample set called D, and searches from the same class for k number of its nearest neighbors called nearest hits H_j. Then it starts searching from each of the different classes for k number of its nearest neighbors called nearest misses M_j.

Following this, it updates the quality estimation $W(A)$ for all attributes A according to Eq. (1):

$$W(A)=W(A)-\sum_{j=1}^{k} diff(A,R_i,H_j)/(mk)+$$

$$\Sigma C\neq class(R_i)[\frac{p(C)}{1-p(class(R_i))}\sum_{j=1}^{k} diff(A,R_i,M_j(C))]/(mk) \quad (1)$$

where m refers to times of repetition, and $diff(A,R_1,R_2)$ signifies the difference between sample R_1 and sample R_2 on feature A, which is measured by Eq. (2):

$$diff(A,R_1,R_2)=\frac{|R_1[A]-R_2[A]|}{Max(A)-Min(A)}. \quad (2)$$

However, although ReliefF is suitable for dealing with high dimensional datasets with a large number of instances, reaching high efficiency in evaluation and excellent performance in noise filtering, it cannot remove redundant features [22].

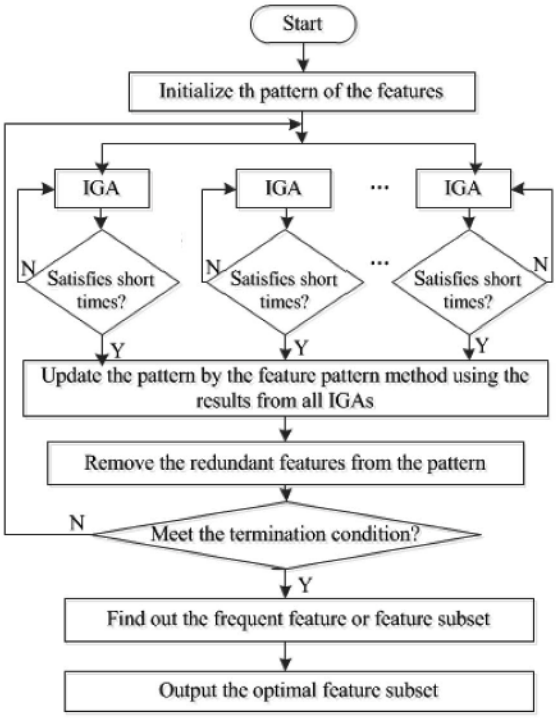


Fig. 1. Method of the feature selection using SPMIGA.

3. Feature selection using multiple improved genetic algorithms with sharing pattern

3.1 The basic flowchart of SPMIGA

First, on the basis of simple genetic algorithm, an improved genetic algorithm is adopted to improve both the search speed and the selecting effect. Then, each population independently executes the IGA operations to search for the optimal feature subsets, in the process of which, each population gets the current best individual after its evolution with short-iterations, and then mining the effective feature pattern according to the best individuals currently obtained by each population. The correlations between the selected features in this feature pattern are further analyzed and redundant features are dropped out. In this way, each population reinitiates individuals according to this feature pattern, and for each individual, if the gene appears in the pattern is assigned "1" value, and otherwise is valued randomly among {0,1}, and then each population continues to perform genetic algorithm operations independently. These procedures are repeated until the algorithm terminates.

In order to further enhance the reliability of the optimal features subsets, and avoid the distraction arising from the occasional features, only those features of high frequency are selected. Fig. 1 illustrated the flowchart of SPMIGA, in which the IGA flowchart is shown in Fig. 2.

3.2 Feature pattern

All the populations select the features simultaneously and, after a fixed number of generations of evolution, each population will communicate and share stage results, so that effective feature pattern is determined and employed to direct the evolution of descendants. Such feature pattern should be composed of important features or feature subsets, which are chosen through the best individual analyses of each population. The whole procedure is shown as follows:

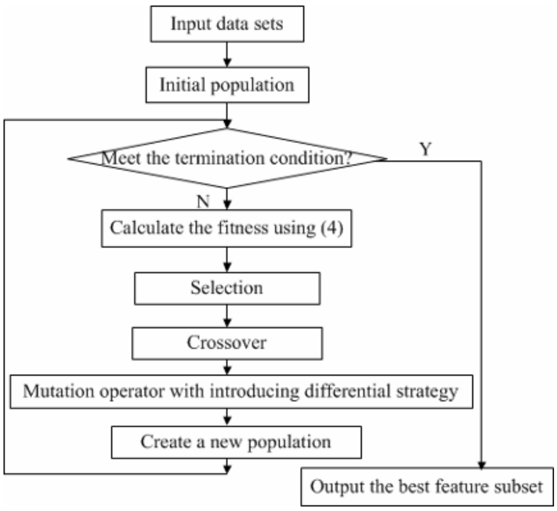


Fig. 2. Flowchart of proposed IGA.

Suppose there are *n* populations, and after *d* generation running, they find their best individuals b_j (j = 1, 2, ... n), in which $b_j = (b_j^1 b_j^2 ... b_j^m)$, *m* refers to the number of individual genes, i.e., the number of features.

Assume that the feature pattern is represented as *pattern* = $p_1 p_2 ... p_m$, where $p_i = p_i + b_j^i$ (i = 1,2,...m) (j = 2,3,..., n), b_j^i stands for the *i*-th feature of the best individual b_j. When p_i value is 0, it means that the current the best individuals of *n* populations do not select the *i*-th feature, which indicates that it is quite probable that this feature is not important. In another word, the higher the p_i value, the greater the importance of this feature.

The initial feature *pattern = 00...0*. After each population has run for *d* generations, update pattern. Each population then initializes the population according to the new pattern: If $p_i \geq t$ ($1 \leq t \leq n$), the *i*-th feature of each population will be assigned "1" value, otherwise, 0 or 1 will be assigned randomly. When $t = 1$, it means the best individual features selected of each population is important; whereas when $t = n$, it indicates that a feature will be important only when all populations take the feature as the best individual. These important features or important feature subsets thus constitute the feature patterns, which not only realize the communication between the populations, but also enable each population to select important features from their own best individuals and use them to guide the evolution of their follow up generation.

3.3 The process of redundant features

Suppose both feature *x* and feature *y* are important features, which contribute greatly to the classification. Since they are important for classification, a population tends select both of them. However, if there is a strong correlation in between and no significant difference could be observed between their combined classification accuracy and that of each single feature, it can be inferred that these two features are redundant. In order to obtain the final high classification accuracy and achieve the relationship between the features of the finally selected feature subsets should be as pure as possible; such redundant features must be removed. In this paper, the processing of redundant features is carried out on the basis of the current feature patterns, which is specified as follows:

(1) Calculating the Pearson correlation coefficient [23] between two features according to Eq. (3), where n is the sample size.

$$\rho_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}} . \tag{3}$$

(2) Calculating the weight of each feature by ReliefF, with high weights assigned to features of high correlation with the classification, and sorting the features according to their weights, the higher the weight, the closer to the top will a feature is listed.

(3) Selecting two features of high correlation coefficient, dropping either one randomly if their weights are very close, that is, setting the value to 0 in the feature pattern, or keeping the feature with the higher weights if their weights are quite different, that is, setting the value to 1 in the feature pattern. In this way, the important features can be retained, and the redundant features can be further removed, so the important and "pure" feature pattern can be used to guide the evolution of offspring.

(4) After the above process of redundant features, new feature models will be obtained.

3.4 IGA

Genetic algorithm was first proposed by professor Holland in 1975. It is a random search method based on the theory of natural selection and survival of the fittest. By selecting operators, crossover operators and mutation operators, it simulates the chromosomal operations when organisms reproduce. The three operators of selection, crossover and mutation are also called the three basic operators of genetic algorithms. Selection operator and crossover operator promoted the evolution of the population but sacrificed the diversity of the population. The mutation operator helps to restore the diversity of the population and prevents the population fall into the local optimum too early by randomly changing a gene or gene on the chromosome. While selecting operator usually adopted roulette selection strategy or tournament selection strategy, the crossover operator mainly falls into one-point crossover, two-point crossover and multiple-point crossover. As to mutation operation, there are mainly uniform mutation and Gaussian mutation. The procedure of the genetic algorithm involves five main factors: The individual coding, the setting of initial population, the design of fitness function (evaluation function), the genetic operators (selection, crossover, mutation and elitist strategy) and, algorithm control parameters. The flowchart of IGA is shown in Fig. 2.

3.4.1 Fitness function

The fitness function is designed based on the principle of minimum intra-class distance and maximum interclass distance [24], and the evaluation unit is feature subset rather than single feature. It is hoped that the selected feature subset can achieve large interclass distance and small intra-class distance.

Therefore, the ratio of the interclass distance to the intra-class distance as the fitness function of a chromosome in the population is defined as

$$f = \frac{\sum_{j=1}^{c} \| \bar{x}^{(j)} - \bar{x} \|^2}{\sum_{j=1}^{c} \frac{1}{n_j - 1} \sum_{k=1}^{n_j} \| x_k^{(j)} - \bar{x}^{(j)} \|^2} \tag{4}$$

where $\bar{x}^{(j)}$ is the mean vector of the feature subset in class j; $\bar{x}$ is the mean vector of the feature subset in the whole samples; $x_k^{(j)}$ is the kth sample vector of the jth class; n_j denotes the number of samples of class j; c represents the number of classes.

3.4.2 Differential mutation operator

In the proposed SPMIGA algorithm, we employ the differential mutation operator [25]. Each new individual is generated by a linear combination of multiple individuals from parent population, rather than the traditional cross operation with the single individual from parent population in GA. The distance obtained from these two parent individuals is then used to determine how many genes to be mutated, and then the gene of the offspring h_i is computed according to Eq. (8). This mutation operator acts to maintain the diversity and improve the efficiency of the algorithm, which is given as follows:

(1) Selecting three different individuals randomly from parent population, as $r1, r2, r3$, respectively.

(2) Computing the distance between $r2$ and $r3$ according to Eq. (5):

$$dist = \sqrt{\sum_{i=1}^{NVARS} (r2^{(i)} - r3^{(i)})^2} \tag{5}$$

where $rj^{(i)}$ represents the ith gene of individual $rj (j = 1,2,3)$; $NVARS$ denotes the number of genes.

(3) Determining how many genes the individual needs to be mutated according to Eq. (6), where p is a random value uni-

Table 1. Basic parameters of bearing 6206. [mm]

Bearing designation	Thickness	Outer race diameter	Inner race diameter	Roller diameter	Pitch diameter
6206	16	62	30	9.5	46

Table 2. Sample data sets of rolling bearing fault diagnosis.

Dataset	Number of features	Number of total samples	Number of the training samples	Number of the testing samples
1800 CV	13	474	315	159
2000 CV	13	475	315	160
2400 CV	13	496	330	166
1800 CH	13	474	315	159
2000 CH	13	475	315	160
2400 CH	13	496	330	166

formly generated within the interval [0, 1].

$$dm = \begin{cases} (\text{int})dist+1, & if\ \ p<(dist-(\text{int})dist) \\ (\text{int})dist, & otherwise. \end{cases} \tag{6}$$

(4) The new gene value after mutation operator: Randomly select a gene i, whose value is expressed as h_i. Then, calculate their combination of the three individuals $r1, r2, r3$ selected and assign the value to h_i according to Eq. (7):

$$h_i = r1^{(i)} + F*(r2^{(i)} - r3^{(i)}) \tag{7}$$

where F is a scaling factor for the difference vector, and generally is set to a constant in the range [0, 2]. Experience shows that if F is too small, it will result in premature convergence, while if F is too large, the convergence speed will be obviously decreased. In this essay, F is set to 1.0.

As $r1, r2, r3$ is 0 or 1, it is easy to see that there are eight kinds of different combinations of the mutually different $r1, r2$ and $r3$ (the coding we adopted is the binary gene model, in which a gene represents a feature), and according to Eq. (7), 6 of the different kinds of combination of the three individuals can achieve binary number 0 or 1, while the other two combinations (001 and 110) were -1 and 2, respectively. Hence, h_i in this paper is defined as follows:

$$h_i = \begin{cases} h_i, & if\ \ h_i = 0\ \ or\ h_i = 1 \\ 0, & if\ \ \dfrac{1}{1+e^{-h_i}} < rand(0,1) \\ 1, & otherwise. \end{cases} \tag{8}$$

4. Selection of rolling bearing fault features

4.1 Introduction to rolling bearing failure simulation experiment

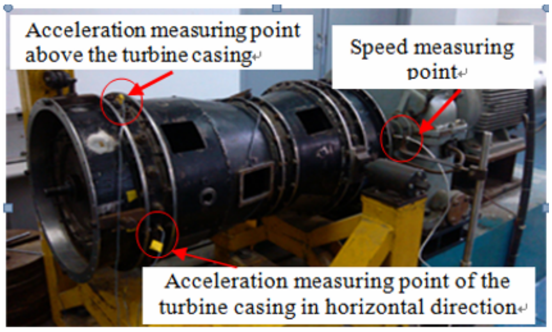


Fig. 3. Aero-engine rotor test.

In this study, a rotor test rig of the aero-engine, which is designed by the Shenyang Aero-engine Design Institute of China, was used to carry out the experiments [26]. The acceleration sensors were arranged above the tester vertical and horizontal direction of the tester to obtain the vibration acceleration signal of the casing. The vibration signal was collected by the NI USB9234 data acquisition device; the acceleration sensor signal is B&K 4805 and the sampling frequency is 10.24 kHz. The tester is 6206 type rolling bearing. The bearing parameter is shown in Table 1, and the tester is shown in Fig. 3. Three sets of fault simulation experiments were carried out at the aero-engine rotating speeds 1800 rpm, 2000 rpm and 2400 rpm. Each experiment produces data out of four situations: Healthy, outer ring fault, inner ring fault, and rolling ball fault (as shown in Fig. 4). There are two measuring points at each speed, where CV is the vertical top of the turbine box and CH is the horizontal direction of the turbine casing.

13 dimensionless characteristic were obtained by time domain, frequency domain and time-frequency analysis [27]. They are the skewness S1, wave factor S2, impact indicator S3, peak indicator S4, kurtosis S5, margin indicator S6, center of gravity frequency S7, mean square frequency S8, frequency variance S9, inner ring frequency envelope spectrum feature S10, outer ring frequency envelope spectrum feature S11, rolling element frequency envelope spectrum feature S12, and the cage frequency envelope spectrum feature S13, respectively. The fault sample data, training samples and test samples are shown in Table 2, where "1800 CH" indicates the

(a) Outer ring fault

(b) Inner ring fault

(c) Rolling ball fault

Fig. 4. Bearing 6206 after fault processing.

sample data obtained in the horizontal direction of the turbine casing at the aero-engine rotating speed 1800 rpm. Other datasets are similarly named.

4.2 Parameter setting

The genetic parameters of IGA and SGA in this paper are set as follows: binary encoding, roulette selection, single point crossover, population size of 50, each algorithm being set to run 50 times, each time running 50 generations, the crossover probability being 0.7, mutation probability being 0.05. While SGA employs the uniform mutation, SPMSGA (multiple simple genetic algorithms with sharing the pattern) and SPMIGA employs five populations, with each population being processed by SGA or IGA for 50 generations and each population communicate with each other every 10 generations.

4.3 Experiment results and analyses

4.3.1 Rate of convergence

Since IGA is based on SGA, and it is to improve the search speed and the performance of the algorithm, and all the populations in SPMIGA algorithm use the IGA to select the features. Therefore, the convergence rate of IGA and SGA is compared below. Meanwhile, in order to test the algorithms' convergence rates in rolling bearing fault feature selection, 1800 CV data set in Table 2 is chosen for sample verification, for, as is shown in Fig. 6, the data set classification accuracy of the two algorithms is the same (both have selected S8, S9, S11 as the optimal feature subset). Therefore, the comparison of convergence rates is based on the fact that both have achieved the same accuracy. The results of convergence rate comparison of IGA and SGA are shown in Fig. 5.

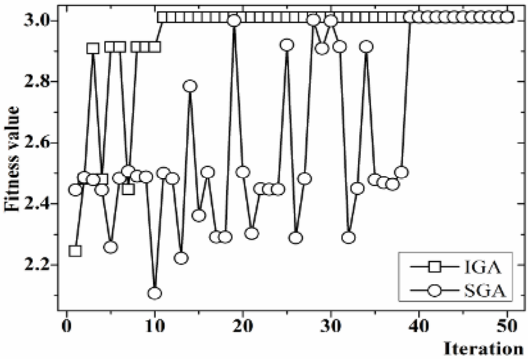


Fig. 5. Convergence graphs of IGA and SGA.

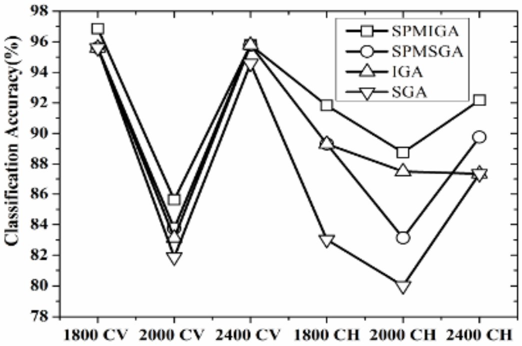


Fig. 6. Classification accuracy by 4 genetic algorithms.

From Fig. 5, it can be observed that the convergence rate of the IGA is much superior to that of SGA. While it takes 10 generation running for IGA to find the optimal feature subset, SGA has to run 40 generations to reach the same result. This shows that the differential strategy adopted by IGA to improve the mutation operator is effective. Through linear combination of three different parent individuals, the improved mutation operator generates a new individual in contrast to the single parent chromosome mutation techniques of SGA, which is beneficial for the new individual to inherit more excellent genes from paternal individuals.

4.3.2 The classification accuracy of the best feature subset

This paper takes 13 features of rolling bearing faults from Table 2 and employs SGA, IGA, SPMSGA and SPMIGA to select the rolling bearing faults respectively. Then it uses Weka software to identify and classify the features select by each algorithm. Their classification accuracy is verified by

J48 10 fold cross validation, where, J48 is a decision tree algorithm based on C4.5 [28] algorithm. The final result is shown in Fig. 6.

First, compare SPMSGA, IGA and SGA to verify the effect of mutation operator and the sharing pattern with multi-population. From Fig. 6, it can be seen that classification accuracy of IGA is superior to that of SGA, with 2 equal results and 4 better ones, which means that IGA are more conducive and accurate in finding the feature subset. Meanwhile, from the comparison of SPMSGA and SGA, it also can be seen that classification accuracy of SPMSGA is superior to that of SGA, with 2 equal results and 4 better ones; therefore, the strategy of the sharing pattern with multiple populations is effective to improve the performance of the algorithm. As for SPMIGA, because it is based on IGA and has adopted multiple populations feature selection, its findings are all superior to that of SGA. Compared with IGA and SPMSGA, SPMIGA is also more effective, with 1 equal and 5 better result. This indicates that not only does SPMIGA bear the advantage of IGA; it also turns out to be more efficient in finding the feature subset because it employs multiple population, in which all the populations communicate on a regular basis and share the features of their best individuals. The shared feature pattern is then passed on to offspring and makes the descendants evolve into faster and better performers and thus enhance the efficiency of feature selection.

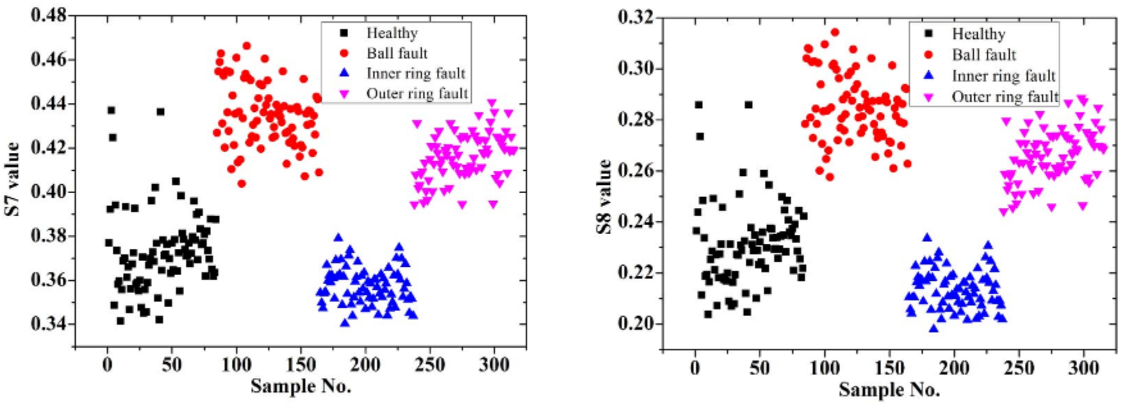
4.4 Feature correlation analysis

The feature correlation analysis is done within 1800CH data sets shown in Table 2. First, Pearson correlation coefficients are calculated pairwise. As is shown in Table 3, strong correlations can be observed in some features of 1800CH data sets, such as S7 and S8, whose Pearson correlation coefficient reaches 0.99. Their class diagram is shown in Fig. 7. Their confusion matrix and classification accuracy as shown in Table 4.

4.4.1 Pearson correlation coefficients

Table 3. Pearson correlation coefficients in 1800 CH data sets.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
S1	1.00	0.86	0.82	0.86	0.98	0.87	0.59	0.59	0.41	-0.15	-0.09	0.00	-0.56
S2		1.00	0.66	0.74	0.76	0.82	0.45	0.45	0.30	-0.07	-0.02	0.03	-0.49
S3			1.00	**0.99**	0.80	0.96	0.55	0.54	0.36	-0.10	-0.03	0.02	-0.51
S4				1.00	0.83	**0.99**	0.56	0.55	0.37	-0.11	-0.03	0.02	-0.53
S5					1.00	0.82	0.57	0.57	0.40	-0.18	-0.10	-0.03	-0.53
S6						1.00	0.54	0.53	0.36	-0.10	-0.03	0.02	-0.53
S7							1.00	**0.99**	0.86	-0.06	0.20	0.28	-0.54
S8								1.00	0.88	-0.10	0.18	0.26	-0.55
S9									1.00	-0.28	0.09	0.16	-0.44
S10										1.00	0.65	0.57	0.29
S11											1.00	0.74	0.21
S12												1.00	0.14
S13													1.00



(a) Scatter plots of feature S7

(b) Scatter plots of feature S8

Fig. 7. Scatter plots of original.

Table 4. Confusion matrix and classification accuracy.

(a) Feature S7

		Predicted class			
		Healthy	Ball fault	Inner ring fault	Outer ring fault
Actual class	Healthy	45	2	30	7
	Ball fault	0	53	0	27
	Inner ring fault	14	0	59	0
	Outer ring fault	0	12	0	66
Classification accuracy				70.79 %	

(b) Feature S8

		Predicted class			
		Healthy	Ball fault	Inner ring fault	Outer ring fault
Actual class	Healthy	42	2	25	14
	Ball fault	0	58	0	22
	Inner ring fault	14	0	59	0
	Outer ring fault	1	16	0	61
Classification accuracy				70.16 %	

(c) Feature S7 & Feature S8

		Predicted class			
		Healthy	Ball fault	Inner ring fault	Outer ring fault
Actual class	Healthy	57	2	18	7
	Ball fault	0	58	0	22
	Inner ring fault	15	0	58	0
	Outer ring fault	2	18	0	58
Classification accuracy				73.33 %	

4.4.2 The scatter plots of feature S7 and feature S8 in different bearing states

Fig. 7 shows the scatter plots of feature S7 and feature S8 in different bearing states, namely healthy, outer ring fault, inner ring fault and ball fault. In all cases, it is difficult to completely distinguish the four states by feature S7 or feature S8.

4.4.3 Confusion matrix and classification accuracy

Confusion matrix depicts the classification accuracy of a classifier, which can be calculated by accuracy, true positive rate, false positive rate and other various evaluation indexes. Table 4 shows the confusion matrix and classification accuracy of feature S7, feature S8, features S7 & S8, as is shown respectively in Tables 4(a)-(c).

It can be seen from Table 4 that the classification accuracy of S7 and S8 is respectively 70.79 % and 70.16 %, which indicates that feature S7 or feature S8 have almost the same contribution to feature classification. Table 4(c) also suggests that even when feature S7 & feature S8 work together, the classification accuracy is not quite different from those of feature S7 and feature S8. This means that there is certain redundancy between feature S7 and feature S8. Therefore, in the feature selection, redundancy must be taken into consideration between high Pearson correlation coefficient features and redundancy processing must take its place. This will not only achieve high classification accuracy, but also keep the features distinctive.

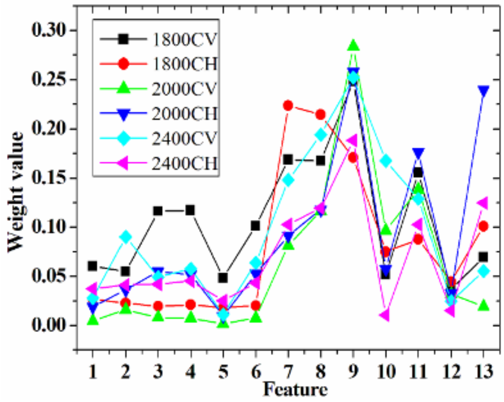


Fig. 8. Feature weight values using ReliefF.

4.4.4 Feature weights

Fig. 8 illustrates the feature weights of 13 rolling bearing fault sample data in Table 2, which are processed with ReliefF algorithm. It can be seen that, with regards to data set of 1800CH, the weight of feature S7 is slightly heavier than that of S8, which indicates that feature S7 contributes to the classification slightly more than feature S8.

After a comprehensive analysis of the scatter plot, Pearson correlation coefficient, confusion matrix, classification accuracy and ReliefF coefficients of the features S7 and feature S8, it is reasonable to remove feature S8 and keep feature S7 in the feature pattern, through which such a periodic result will

Table 5. Selected feature subsets obtained by 4 algorithms.

Dataset	Selected feature subset			
	SPMIGA	ReliefF	CFS	InfoGain
1800 CV	3,7,9	7,8,9	5,7,8,9,10,11,13	7,8,9
2000 CV	2,3,8,9,10,11	7,8,9,10,11,12	1,7,8,9,10,11,12	1,7,8,9,10,11
2400 CV	7,9,10	8,9,10	1,8,9,10,13	7,8,9
1800 CH	7,9,11	7,8,9	5,7,8,9,10,11,12,13	7,8,9
2000 CH	1,7,9,11	8,9,11,13	6,8,9,11,12,13	8,9,11,13
2400 CH	1,7,9,11	7,8,9,13	5,8,9,10,11,13	5,7,8,9

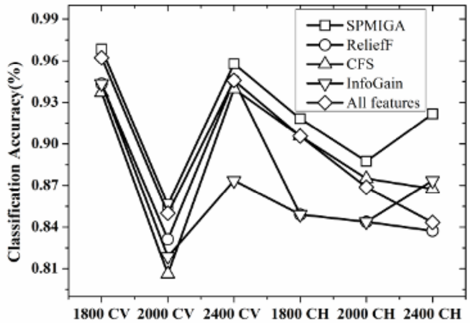


Fig. 9. Classification accuracy by 5 algorithms.

be passed to the descendants.

4.5 Comparison of the classification accuracy of SPMIGA and that of other algorithms

In this paper, SPMIGA employs multiple population feature selection, which introduces a regular feature communication and redundancy processing. In order to prove its validity in feature selection, it is necessary to compare its feature selection results with other algorithms and all the features. Considering the validity of the comparison, as ReliefF and InfoGain [29] ranking the feature by their weight, all the features selected by each algorithm will be ranked in the same way and equal number of top features are taken for comparison. The results are shown in Fig. 9 and Table 5, respectively.

As can be seen from Fig. 9, SPMIGA performs more efficiently in terms of classification accuracy than ReliefF, CFS [30] (correlation-based feature selection) and InfoGain, with all win advantage over CFS and InfoGain in the 6 data sets and only one slightly inferior result to that of ReliefF. Take 1800CH data set for example, as there is a strong correlation between feature S7 and feature S8, redundancy must be taken into consideration in the feature selection. For this reason, classification accuracy of SPMIGA is the highest as compare with the other three algorithms.

In addition, the classification accuracy of SPMIGA is compared with that of all features, and SPMIGA is dominant on the 6 data sets. In other words, SPMIGA obtains higher classification accuracy with fewer features, which also further illustrates the necessity and importance of dealing with features that have strong correlation or redundancy. Therefore, this result proves that SPMIGA put forward in this paper in efficient and valid in enhancing the classification accuracy in feature selection and, the features selected are distinctive.

5. Conclusions

(1) This paper proposes a sharing pattern multi-population genetic algorithm for feature selection, which was based on the simple genetic algorithm, adopted a multiple-population improved genetic algorithm and, improved the speed and effect of algorithm search result. Meanwhile, SPMIGA employs multiple population mechanism in which all populations share and communicate their best individuals on a regular basis in the process of feature selection and thus direct all populations evolve with the help of feature patterns. During this process, redundancy must be taken into consideration between high Pearson correlation coefficient features and redundancy processing must take its place.

(2) In order to prove the validity of SPMIGA in feature selection, a comparison research is carried out based on the experimental data of the rolling bearing faults of an aero-engine, which involves SPMIGA, ReliefF, CFS and InfoGain. The result shows that the classification accuracy of SPMIGA is the highest as compare with the other three algorithms and thus proves that SPMIGA put forward in this paper in efficient and valid in enhancing the classification accuracy in feature selection.

Acknowledgments

This research is sponsored by the National Natural Science Foundation of China (No. 51675263); by Jiangsu Postgraduate Research Innovation Projects (No. KYLX15_0315), by the Natural Science Foundation of Guangdong Food and Drug Vocational College (No. 2016YZ008).

References

[1] O. Gustafsson and T. Tallian, Detection of damage in assembled rolling element bearings, *A S L E Transactions*, 5 (1) (1962) 197-209.

[2] P. G. Wheeler, Bearing analysis keeps downtime down, *Plant Engineering*, 25 (1968) 87-89.

[3] R. Martin, Detection of ball bearing malfunctions, *Instruments and Control Systems*, 12 (1970) 79-82.

[4] X. Zhao and B. Ye, Study on clear extraction of amplitude

spectrum of vibrating signal for a rolling bearing by art-2 and its classification effect, *Journal of Vibration & Shock*, 26 (1) (2007) 139-143+150.

[5] B. E. Parker Jr. et al., Diagnostics using statistical change detection in the bispectrum domain, *Mechanical Systems and Signal Processing*, 14 (4) (2000) 561-570.

[6] Z. Y. Rui, L. Y. Xu and G. M. Li, Fault detection of roller bearing based on cepstrum, *Bearing*, 1 (2007) 35-37.

[7] J. T. Yang, J. J. Chen and Z. P. Zeng, Extracting fault features using higher order spectra for rotating machinery, *Journal of Vibration Engineering*, 14 (1) (2001) 13-17.

[8] P. W. Tse, Y. H. Peng, R. Yam and R. Yam, Wavelet analysis and envelope detection for rolling element bearing fault diagnosis-their effectiveness and flexibilities, *Journal of Vibration & Acoustics*, 123 (3) (2001) 303-310.

[9] G. Chen, Feature extraction and intelligent diagnosis for ball bearing early faults, *Acta Aeronautica ET Astronautica Sinica*, 30 (2) (2009) 362-367.

[10] J. S. Cheng, D. J. Yu and Y. Yang, Fault diagnosis of roller bearings based on EM D and SVM, *Journal of Aerospace Power*, 21 (3) (2006) 575-580.

[11] J. Antoni and R. B. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, *Mechanical Systems and Signal Processing,* 20 (2006) 308-331.

[12] F. Cong et al., Short-time matrix series based singular value decomposition for rolling bearing fault diagnosis, *Mechanical Systems & Signal Processing*, 34 (1-2) (2013) 218-230.

[13] A. Akram, N. Rami and A. A. Ahmed, Enhancing the diversity of genetic algorithm for improved feature selection, *2010 IEEE International Conference on Systems Man and Cybernetics (SMC),* Piscataway, NJ: IEEE Press (2011) 1325-1331.

[14] Q. W. Yang et al., Improving genetic algorithms by using logic operation, *Control and Decision*, 15 (4) (2000) 510-512.

[15] Y. Li, S. Zhang and X. Zeng, Research of multi-population agent genetic algorithm for feature selection, *Expert Systems with Applications*, 36 (9) (2009) 11570-11581.

[16] D. Dutta, P. Dutta and J. Sil, Simultaneous feature selection and clustering for categorical features using multi objective genetic algorithm, *International Conference on Hybrid Intelligent Systems, IEEE* (2012) 191-196.

[17] P. Chen, T. Toyota and Z. He, Automated function generation of symptom parameters and application to fault diagnosis of machinery under variable operating conditions, *Systems Man & Cybernetics Part A Systems & Humans IEEE Transaction on*, 31 (6) (2001) 775-781.

[18] M. Kang et al., Time-varying and multi-resolution envelope analysis and discriminative feature analysis for bearing fault diagnosis, *Industrial Electronics, IEEE Transactions on*, 62 (12) (2015) 7749-7761.

[19] M. M. Ettefagh, M. Ghaemi and M. Y. Asr, Bearing fault diagnosis using hybrid genetic algorithm K-means clustering, *Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium on. IEEE* (2014) 84-89.

[20] M. Mitchell, *An introduction to genetic algorithms*, Cambridge, MA: MIT Press (1996) 2, ISBN 9780585030944.

[21] I. Kononenko, Estimation attributes: Analysis and extensions of ReliefF, *Proceedings of the 1994 European Conference on Machine Learning,* Catania, Italy: Springer Verlag (1994) 171-182.

[22] L. X. Zhang et al., Combination feature selection based on Relief, *Journal of Fudan University (Natural Science)*, 43 (5) (2004) 893-898.

[23] K. Pearson, Note on regression and inheritance in the case of two parents, *Proceedings of the Royal Society of London*, 58 (2006) 240-242.

[24] X. G. Zhang, *Pattern recognition*, Third edition, Beijing: Tsinghua University Press (2010) 146-148.

[25] X. Y. Guan, G. Chen and T. Lin, Feature selection method based on differential evolution and genetic algorithm with multi-criteria evaluation and its applications, *Acta Aeronautica et Astronautica Sinica*, 37 (11) (2016) 3455-3465.

[26] G. Chen et al., Sensitivity analysis of fault diagnosis of aeroengine rolling bearing based on vibration signal measured on casing, *Journal of Aerospace Power*, 29 (12) (2014) 2874-2884.

[27] H. B. Mei, *Rolling bearing vibration monitoring and diagnosis*, Beijing: China Machine Press (1995).

[28] I. H. Witten, E. Frank and M. A. Hall, *Data mining: Practical machine learning tools and techniques*, Third Edition, Beijing: China Machine Press (2005).

[29] L. Yu and H. Liu, Efficient feature selection via analysis of relevance and redundancy, *Journal of Machine Learning Research*, 5 (2004) 1205-1224.

[30] M. A. Hall, Correlation-based feature selection for discrete and numeric class machine learning, *Seventeenth International Conference on Machine Learning,* Morgan Kaufmann Publishers Inc (2000) 359-366.

X. Y. Guan received a master degree in the School of Software from the Sun Yat-sen University, Guangzhou, P. R. China, in 2008. Now she is a student in the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, P. R. China. Her current research interests include genetic algorithm, pattern recognition and machine learning, and their application in bearing fault diagnosis.

G. Chen received a Ph.D. degree in the School of Mechanical Engineering from the Southwest Jiaotong University, Chengdu, P. R. China, in 2000. Now he works at the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, P. R. China. His current research interests include the whole aero-engine vibration, rotor-bearing dynamics, rotating-machine fault diagnosis, pattern recognition and machine learning, signal analysis and processing.