Measurement 173 (2021) 108654

Contents lists available at ScienceDirect

Measurement

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

Fault diagnosis of planetary gearbox using multi-criteria feature selection and heterogeneous ensemble learning classification

Zirui Wang, Haian Huang, Youren Wang

College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, Jiangsu Province, PR China

ARTICLE INFO

Keywords:
Planetary gearbox
fault diagnosis
vibration signal
multi-criteria feature selection
multi-objective evolutionary algorithm
Dezert-Smarandache theory
heterogeneous ensemble learning

ABSTRACT

Fault feature selection and early fault diagnosis for planetary gearbox are important tasks and have been widely investigate. This paper proposes a novel fault diagnosis scheme for planetary gearbox using multi-criteria fault feature selection and heterogeneous ensemble learning classification. Vibration signals collected by the acceleration sensors are imported for fault diagnosis of planetary gearbox. High dimension fault features are extracted by analyzing the vibration signals in time domain, frequency domain and time-frequency domain. The criteria for selecting lower dimension optimal fault features of planetary gearbox are designed, and the mathematic model for fault feature selection with multi-criteria is established. After that, a new feature selection method using Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) is applied to obtain diverse lower dimension quasi optimal fault feature subsets. Then each quasi optimal fault feature subset is transferred to a base classifier for primary fault diagnosis. Those base classifications are performed by support vector machine and sparse Bayesian extreme learning machine respectively. Dezert-Smarandache rules are applied for classifier-level fusion to achieve and evaluate overall accuracy of the fault diagnosis for planetary gearbox. Experimental results state that the proposed method constantly gets diverse lower dimension quasi optimal fault features smoothly, and significantly improves the accuracy and robustness of fault diagnosis.

1. Introduction

Planetary gearbox plays an important role in wind turbine generator system, hydraulic generator sets, helicopters and other systems with mechanical power transmission. It often has a high failure rate as the working conditions of planetary gearbox is complex and variable, such as relatively high or low temperature, time-varying speed and dynamic load [1,2]. At present, most fault diagnosis methods for planetary gearbox obey the following process. Firstly, they extract fault features from vibration signals [3]. Then pattern recognition or machine learning algorithm are employed to identify the fault modes. The general features of vibration signals include time domain feature such as root mean square (RMS), kurtosis, skewness, NA4, frequency domain feature such as central frequency or time-frequency spectrum. Moreover, several kinds of new fault features are developed recently. Lei [4] proposed two fault features of the planetary gearbox vibration signals, which are accumulative amplitudes of carrier orders and the energy ratio of differential spectrum. Li [5] proposed a novel fault diagnosis method based on modified multi-scale symbolic dynamic entropy, which can quantify the regularity of time series, simplify the computation process and clearly detect signal dynamic variation.

For machine learning-based fault diagnosis methods, a sizable amount of fault features could be extracted from vibration signals which are gradually processed to obtain condition parameters in time domain, frequency domain and time-frequency domain [6,7]. Contrary to theoretical expectations, an increase numbering of fault features does not improve the discriminating power of the machine learning model for fault diagnosis [8]. Instead they result in over-fitting problems due to the redundant features and irrelevant fault features. Furthermore, the size of fault features has significant impact on the computational complexity of the machine learning algorithms. These problems are typically related to the fault feature dimensionality. Hence, feature selection methods are widely applied to select the most relevant features and improve the accuracy of machine learning fault diagnostic model. In recent years, a variety of fault feature selection techniques for planetary gearbox fault diagnosis have been developed to reduce the feature size, remove irrelevant features and redundant features from the original feature set. Z. Liu [9] proposed a multi-criterion fusion framework for planetary gearbox fault feature selection which took two aspects into account, effectiveness and correlation. In order to solve the non-

E-mail address: 781530780@163.com (Z. Wang).

https://doi.org/10.1016/j.measurement.2020.108654
Received 16 August 2017; Received in revised form 10 September 2020; Accepted 21 October 2020
Available online 1 November 2020

sensitive feature interference in manifold learning fault diagnosis, Baoping Tang [10] proposed a multi-criteria fusion feature selection method according to Dezert-Smarandache Theory, this method evaluates fault features from three aspects, distance, correlation and information entropy. Pacheco [11] proposed a new unsupervised feature selection algorithm using attribute clustering and rough set theory. Attribute clustering consist of two main parts, distanced-based classification and prototype-based clustering, to group similar features. After that rough set theory is used to compute correlation between similar features. Additionally, this feature selection algorithm has on-line training ability and its effectiveness has been verified through evaluating the level of faults in gears and bearings. However, following disadvantages still exist in these fault feature selection methods. Non-sensitive features cannot be eliminated completely because those multi-criteria fusion method is only the simple combination of multiple criteria. On the other hand, those optimal features are selected under their own optimization criterion. In fact, there may be multiple sets of diverse quasi optimal fault features (for chosen) which can increase the fault classification accuracy.

The machine learning-based classification algorithms are research hotspot and have been widely applied to intelligent fault diagnosis of planetary gearbox in recent years, such as support vector machine (SVM), BP neural network, etc. Meanwhile, fault diagnosis approaches based on information fusion at fault features' level and classifier decision level also receives attention. C. Li [12] combined vibration signal and acoustic emission signal and extracted the wavelet packet features simultaneously. They employed two Boltzmann machines to form the deep random forest for information fusion. Xuedong Wang [13] proposed a fault diagnosis method which conducts feature reduction based on Kernel Supervised Locality Preserving Projection (KSLPP), then they use Relief-F feature weighting K-Nearest Neighbor Classifier on the reduced feature subsets for fault diagnosis of rotating machinery. This can enhance the low recognition rate of high-dimensional fault features for rotating machines. J.Yu [14] proposed a new manifold learning algorithm of Joint Global and Local/Nonlocal Discriminant Analysis (GLNDA). GLNDA is used to carry feature selection on each sub-sample constructed by Boostrap sampling. The discrete Particle Swarm Optimization (PSO) algorithm is applied for selecting the best base model. Result is decided by the maximum-voting method. In all, the ensemble learning fusion based fault diagnosis method complete the information and utilize parameter sharing strategy to achieve better fault diagnosis performance compared with single classifier. This is the potential prospect of fault diagnosis technology.

This paper proposed a novel fault diagnosis method for planetary gearbox depending on diverse quasi optimal fault feature fusion and heterogeneous ensemble learning, in order to solve the disadvantages in those fault feature selection methods and improve the performance of early fault diagnosis. Fault feature selection criteria have been designed and a mathematical model has been established. MOEA/D evolutionary algorithm [15] has been used to obtain diverse quasi optimal feature subsets. Heterogeneous ensemble learning is employed for information-fusion-based fault diagnosis of planetary gearbox to improve fault diagnosis performance and its robustness against noise.

The remaining part of this paper are organized as follows. Part 2 introduces the multi-criteria fault feature selection including its mathematic model and multi-objective evolutionary algorithm. Part 3 proposes a fault diagnosis method using quasi optimal fault features and heterogeneous ensemble learning classification. Part 4 describes the experiment and results, which identify the capability of proposed method. Conclusions are outlined in Part 5.

2. Method of multi-criteria fault feature selection

2.1. Criteria of fault feature optimization

Before designing a fault feature selection criterion, the original fault features of planetary gearbox must be acquired and extracted. Here accelerometer sensors are equipped to gather vibration signal from planetary gearbox. Next step is extracting the signal's time domain features, frequency domain features and wavelet domain features. Time domain features are the simplest way for monitoring rotating machine operation conditions, frequency domain features indicate signal property and wavelet domain features enable multiscale time-frequency transient analysis on the signal. The establishment of high-dimensional fault features brings a comprehensive evaluation on the operating condition of gearbox. The extracted 39 features are described as follows:

1) Time domain features [16]:maximal value x_1, minimum value x_2, peak value x_3, peak to peak value x_4, mean value x_5, average absolute value x_6, square mean root x_7, variance x_8, standard deviation x_9, root mean square x_{10}, kurtosis x_{11}, skewness x_{12}, shape factor x_{13}, crest factor x_{14}, impulse factor x_{15}, margin factor x_{16}, clearance factor x_{16}.
2) Frequency domain features [17]: mean frequency x_{18}, central frequency x_{19}, frequency RMS value x_{20}, frequency standard deviation x_{21}.
3) Wavelet domain features (after db6 wavelet transform, with the decomposition level set to 3) [18,19]: wavelet energy ratio x_{22}~x_{29}, wavelet energy entropy x_{30}, wavelet feature scale entropy x_{31}~x_{38}, wavelet singular entropy x_{39}.

The objective of fault feature selection is to improve the fault classification performance by avoiding the interference of redundant and irrelevant fault features. Assume M fault features extracted from vibration signals consist a high-dimensional fault feature vector or fault feature set$\mathrm{x}_F = (x_{F(1)}, x_{F(2)}, \cdots, x_{F(M)})$, M=39, $x_{F(i)}$is the No. i fault feature apply multi-criteria to select optimal fault feature subset from fault feature setx_F.Suppose that data set D is combined of fault feature set sample space X and modey, sample space X is composed of N fault feature set samples, y includes normal mode and fault mode of planetary gearbox.

$$D_{N\times 40} = \{\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_{39}, \mathrm{y}\} \tag{1}$$

Besides, N is the number of fault feature sets.$\mathrm{x}_m = (x_{m,1}, x_{m,2}, \cdots, x_{m,N})^T$represents the N sample values(include normal samples and fault samples) on No. m feature, $\mathrm{x}_m \in \mathrm{X}$, $m = 1,2,\cdots,39$. $x_{m,n}$ is the No. n sample value on the No. m feature. And $\mathrm{y} = (y_1, y_2, \cdots, y_N)^T$,$y_n$is the mode label of the No. n sample.$y_n \in \{1, 2, \cdots, C\}$, assume that the number of all modes to be classified is C.

Fault feature selection process aims to select K features from the original M-dimension fault feature set x_F,and forming the lower-dimensional fault feature subset $\mathrm{x}_Q = (x_{Q(1)}, x_{Q(2)}, \cdots, x_{Q(K)})$.$S_{N\times K}$ is a fault feature subset in lower-dimension fault feature space $\mathrm{x}_Q$acquired from high-dimensional fault feature set sample space X, $0 < K < M$.

To select the effective and optimal fault feature subset, 4 fault feature optimization criteria are listed.

1) Maximum relevance criterion, which evaluate the relevance between fault feature and fault mode. That means, from the prospective of clustering, it does benefit to fault classification when the fault feature is much closer to the cluster center of its fault mode.

$$f_1(S) = \sum_{i=1}^{K} SU(\mathrm{x}_i, \mathrm{y}) \tag{2}$$

$S$is$S_{N\times K}$,x_i is composed of the N sample values from the No. i dimension fault feature, $i = 1, 2, \cdots, K$.

SU is symmetric uncertain information, its calculation formula is listed.

$$SU(\mathrm{u}, \mathrm{v}) = \frac{2I(\mathrm{u}, \mathrm{v})}{H(\mathrm{u}) + H(\mathrm{v})} \tag{3}$$

$H(\mathrm{u})$ is the information entropy of vector u, $I(\mathrm{u}, \mathrm{v})$ is the mutual information of vector u and v, $\mathrm{u} = (u_1, u_2, \cdots, u_N)^T$, $\mathrm{v} = (v_1, v_2, \cdots, v_N)^T$.

$$H(\mathrm{u}) = -\sum_{n=1}^{N} p(u_n) log p(u_n) \tag{4}$$

$p(u_n)$ is the probability value of variable u_n.

$$I(\mathrm{u}, \mathrm{v}) = H(\mathrm{u}) + H(\mathrm{v}) - H(\mathrm{u}, \mathrm{v}) \tag{5}$$

$H(\mathrm{u}, \mathrm{v})$ is the joint entropy of vector u and v.

$$H(\mathrm{u}, \mathrm{v}) = -\sum_{n=1}^{N} \sum_{n'=1}^{N} p(u_n, v_{n'}) log p(u_n, v_{n'}) \tag{6}$$

Besides,$p(u_n, v_{n'})$ is the Joint probability, $p(u_n, v_{n'}) = p(u_n)p(v_{n'}|u_n) = p(v_{n'})p(u_n|v_{n'})$, $0 \leqslant f_1(S) \leqslant 1$.

2) Minimum redundancy criterion, which evaluate the relevance between fault features. Low redundancy indicates the high relevance between two fault features. However, none of the fault features can completely replace another one even if the redundancy is extremely small. Fault feature optimization requires the redundancy to be as small as possible, its calculation formula listed here.

$$f_2(S) = \sum_{k=1, k<k'}^{K} SU(\mathrm{x}_k, \mathrm{x}_{k'}) \tag{7}$$

Where x_k and $\mathrm{x}_{k'}$ are the No. k feature and No. k' feature in fault feature subset sample space S, the calculation formula of SU function is equation (3), $0 \leqslant f_2(S) \leqslant 1$.

3) Minimum feature dimension criterion, which requires the dimension of selected fault feature subsets to be as low as possible after basic fault diagnosis performance is satisfied.

$$f_3'(S) = |K| \tag{8}$$

$0 < f_3'(S) < M$, here for calculation, normalize $f_3(S) = \frac{f_3'(S)}{M}$ between 0-1.

4) Best fault diagnosis rate criterion, which requires the highest successful rate of fault mode classification. Under ideal condition this could be 100%.

$$f_4(S) = \frac{Q}{N} \tag{9}$$

Q is the number of samples which are correctly diagnosed in total N testing samples by using selected fault feature subset. $0 \leqslant f_4(S) \leqslant 1$.

2.2. Mathematic model for multi-criteria fault feature selection

According to the fault feature selection criteria introduced in part 2.1, the mathematic model for selecting optimal fault feature can be described as follows:

Objective function:

$$\min F(S) \tag{10}$$

Besides, $F(S) = \{-f_1(S), f_2(S), f_3(S), -f_4(S)\}$.

Constraint to:

$$\begin{cases} f_1(S), f_2(S), f_3(S), f_4(S) \in (0, 1) \\ S \in \mathrm{X} \end{cases} \tag{11}$$

As the four criteria for selecting optimal fault features are conflicting, the multi-criteria fault feature selection can actually be modeled as multi-objective optimization problems (MOPs). A Pareto-optimal solution to an MOP is a candidate for the optimal tradeoff. An MOP may have many, even an infinite number of Pareto-optimal solutions. The Pareto set (PS)/Pareto front (PF) is the set of all the Pareto-optimal solutions in the objective space. Without the criterion preference information, all the Pareto-optimal solutions are equally good. The good approximation to the PS and/or the PF is wanted for fault diagnosis of planetary gearbox. Because of the complex interactions among fault features, different feature subsets with similar fault mode discriminating power can exist.

2.3. The algorithm for selecting diverse quasi optimal fault features

Conventional genetic algorithm which generally gets only one optimal fault feature selection solution instead of diverse fault feature subsets are employed to solve equation (10). Moreover, Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) is applied.

MOEA/D is a multi-objective evolutionary algorithm consist of mathematical programing and evolutionary algorithms, which is specially developed for multi-objective optimization problem. MOEA/D use decomposition method to convert multi-objective optimization problem into single objective sub-problem. Then population evolution is completed by the evolutionary operation between adjacent sub-problems, and process regular single objective optimization algorithm (e.g. diversity maintenance strategy of genetic algorithm) for elitism strategy. Compared with regular multi-objective optimization algorithm, it has simple principles, quick evolution speed and high computational efficiency which leads to a fast and efficient way acquiring equally distributed diverse optimal fault feature subsets.

Next, the process and calculation steps of MOEA/D would be discussed.

1) Coding scheme. Decision variable (potential solution) d is designed by real number coding. $\mathrm{d} = (d_1, d_2, \cdots, d_m, \ldots, d_{39})^T$, $d_m \in (0, 1)$. If$d_m < 0.5$,the No. m feature of original fault feature set x_F will be abandoned, and $d_m \geqslant 0.5$means select the No. m feature.
2) Description of feature selection process. First convert 39-dimension decision vector d into binary vector which contains only 0 or 1, then use the binary vector to extract fault feature subset sample space S from high-dimensional fault feature space X. The update of decision vector d indicates the continuous optimization process of fault feature selection.
3) Initialization of algorithm. The initial setup contains the following steps. Set objective number to 4, set the number of original fault feature dimension $M = 39$ and population number $N_p = 100$(in other words, the number of sub-problem is 100). Neighbor number T is set to 20, maximum evolution iteration for stop condition is 100. Moreover, execute random initialization on the weight vector of each sub-problem, then fix neighbor of every populations according to the Euclidean distance between weight vectors.

4) The calculation for fitness function. Fitness function can be calculated by Tchebycheff approach [17,18].

$$\begin{cases} ming^{TE}(S|\lambda, z^*) = \max\limits_{1\leqslant i\leqslant 4}\{\lambda_i|f_i(S) - z_i^*\} \\ s.t.S \in X \end{cases} \tag{12}$$

Whereλ is the weight vector of each population, $\lambda = (\lambda^1, \lambda^2, \cdots, \lambda^{N_p})^T$. S is the feature subset selected by decision vector d in step 2). $f_i(S)$ represents the No.i criteria. The fault diagnosis accuracy$f_4(S)$ is calculated by using Sparse Bayesian Extreme Learning Machine (SBELM) classifier and 10-fold cross-validation method. $z^* = (z_1^*, z_2^*, z_3^*, z_4^*)^T$ is the optimal value under four criterion in the current population evolution, according to equation (10). Set $z_i^* = min\{F_i(S)|S \in X\}, i = 1, 2, 3, 4$.

5) Selection, Crossover and Mutation steps.

Selection: From each population, randomly select two neighbors which differ from their own population and each other, to ensure the exploring ability of the algorithm and the diversity of those solutions.

Crossover: Crossover takes Differential Evolution Operator (DEO). For any population, suppose two selected neighbor vectors are d^{r1} and d^{r2}, its computation equation can be expressed as follows.

$$q_m = \begin{cases} d_m + F' \times (d_m^{r1} - d_m^{r2}) & if\ rand(0,1) \leqslant CR \\ d_m & otherwise \end{cases} \tag{13}$$

Besides, F'andCR are the control parameter in the differential evolution, both set to be 0.5. q is the temporary decision vector after Crossover, $q = (q_1, q_2, \cdots, q_m, \cdots, q_{39})^T$.

Mutation: Here take Gaussian mutation, that is confirming 39 random variables which obey Gaussian distribution$N(0, \delta^2)$ with mean value equal to 0 and variance equal to δ^2=1/39. Then substitute the temporary decision vector q into equation (14) to acquire new decision vector x', and limit its boundary to (0,1).

$$x' = q + N(0, \delta^2) \tag{14}$$

6) Population update.

The generated population is evaluated through comparing the Tchebycheff equation's value of new decision vector and the current neighbor decision vector. Their neighbor decision vector and own decision vector will be updated step by step during the comparison process.

The population is composed of the best solution found so far for each sub-problem. Only the current solutions to its neighboring sub-problems are exploited for optimizing a sub-problem.

Apply feature selection on new vector to get new target function $FV^j_{(1,i+1)}$. If , update the reference point. Otherwise, hold the old reference point. Perform same steps on z_2, z_3, z_4.

The fitness function of population update is acquired through Tchebycheff equation. Since is the generated new decision vector and $FV^j_{(1,i+1)}$ is the new 4 dimensional target function, this new rules uses reference point z, weight vector W, neighbor B and Tchebycheff equation to determine fitness value g^{TE} of new decision vector. Meanwhile, in each population, their 20 neighbors' fitness value g^{TE} can also been calculated by Tchebycheff equation. Population update is realized by comparing the neighbors' fitness value with . If the former is smaller, replace their neighbors' input vector by new decision vector.

The MOEA/D algorithm for selecting multi-criteria fault features works is shown as follows. A flow chart is also listed.

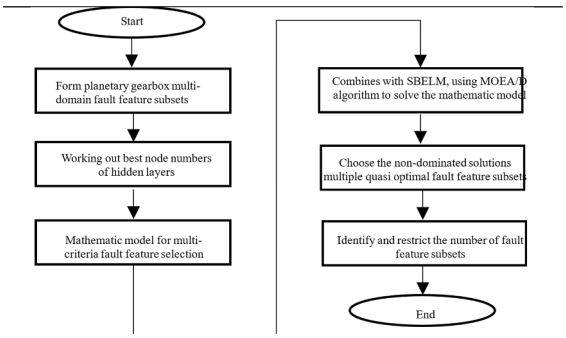


Fig. 1. Flow chart of the proposed fault feature optimization algorithm.

Input: M-dimension fault feature space X, modey

1. Set population number N_p, neighbor number T, maximum iteration times (*Count*) for stopping criterion, differential evolution operator parameter $F'$and CR both to 0.5, Gaussian mutation variance δ^2=1/39.
2. Randomly initialize weight vectors for N_p populations and use Euclidean distance to fix neighbors for every population.
3. For $i = 1, 2, \cdots, Count$
For $j = 1, 2, \cdots N_p$
   3.1 Convert real-number decision vectordinto binary code and get feature subset S_i^j.
   3.2 Calculate the objective function value for selected feature subset S_i^j, and use that value to calculate the minimum criteria value z^* of the whole population.
   3.3 Select two neighbors from current population, execute crossover with differential evolution operator. Then apply Gaussian mutation and limit the boundary of new decision vector x'.
   3.4 Calculate fitness function over new decision vector x' and neighbor decision vector separately, replace the neighbor decision vector with new decision vector x' if neighbor decision vector has better fitness function value. Update z^* at the same time.
   3.5 Break 'for' loop if reaching maximum iterations. Otherwise, continue step 3.
4. Result in diverse fault feature subsets, choose the multiple quasi optimal fault feature subsets with their maximum deviation rate less than 5%.

Output: multiple quasi optimal fault feature subsets (Fig. 1).

3. Fault diagnosis method based on quasi optimal features fusion and heterogeneous ensemble learning classification

In this part, the diverse quasi optimal fault features are taken for information integration. Heterogeneous ensemble learning based fault diagnosis algorithm is raised which boosts the performance of fault diagnosis by three aspects: conducting random re-sample of input data; ensuring the input to base classifier is diverse quasi optimal fault feature subset; using heterogeneous ensemble learning fault classification. Base classifier uses Support Vector Machine (SVM) and Sparse Bayesian Extreme Learning Machine (SBELM). SVM are trained by fault feature subsets which rank in top of their classification accuracy rate, SBELM are trained by the other subsets. The outputs of these base classifiers are fused by Dezert-Smarandache Theory (DSmT) for decision level fusion which enhances its fault diagnosis performance.

3.1. Support vector machine

SVM works on the principle of structural risk minimization, which searches an optimal linear separation plane in the high-dimensional kernel space. SVM has the advantages of high stability, high generalization ability, robustness against noises, and its convergence can be theoretically proved. So SVM is widely applied in the fault classification under small sample cases.

Given a data set D contains N fault feature samples, see equation (15), SVM transform optimal hyperplane problem into dual quadratic programming problem by importing the method of Lagrange multipliers:

$$\begin{cases} maxL(\mu) = \sum_{n=1}^{N}\mu_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N} y_n y_{n'}\mu_n\mu_{n'}(x_n \cdot x_{n'}) \\ s.t. \sum_{n=1}^{N} y_n\mu_n = 0, 0 \leqslant \mu_n \leqslant P \end{cases} \tag{15}$$

P is penalty parameter, designed to control the penalty level of misclassified samples. μ_n is Lagrange multiplier. Optimal hyperplane coefficient can be acquired through solving the optimal solution for μ.

3.2. Sparse Bayesian extreme learning machine

Common Extreme Learning Machine (ELM) has two shortages in operation. First one is that the output weight value of ELM is determined by Moore-Penrose generalized inverse, which easily leads to overfitting. The other one is ELM has precise requirement on the number of nodes in a hidden layer thus it generally requires a large number of nodes. SBELM solved the generalized inverse solution problem occurred in ELM by estimating marginal likelihood distribution during learning phase. It also has the ability to minimize the spare nodes in a hidden layer by setting output hyper-parameters. In all, SBELM has following advantages [20 21]: probability output, high generalization ability, sparsity, accelerated learning ability and so on.

Marginal likelihood distribution can be represented:

$$p(y|\alpha, H) = \int p(y|\beta, H)p(\beta|\alpha)d\beta \tag{16}$$

$p(y|\beta, H)$ is likelihood function which satisfies Bernoulli distribution. $p(\beta|\alpha)$ satisfies 0 mean value Gaussian prior distribution and constrain to hyper-parameter α_j of automatic relevance determination (ARD) [22].

$$p(\beta|\alpha) = \prod_{i=0}^{L}\frac{\alpha_i}{\sqrt{2\pi}}exp(-\frac{\alpha_i\beta_i^2}{2}) \tag{17}$$

β is output weight of the hidden layer. L is the number of nodes in a hidden layer. H is the output matrix of the hidden layer.

3.3. Dezert-Smarandache theory of evidence

The Dezert-Smarandache theory (DSmT) of plausible and paradoxical reasoning proposed by the authors in recent years [23] is an advanced theory based on classic Dempster-Shafer theory of evidence (DST). DSmT specially focus on the fusion of uncertain, highly conflicting and imprecise evidence sources. Thus, it has advances in solving complex static or dynamic fusion problems beyond the limits of DST framework.

DSmT is briefly introduced as follows.

(a) Foundations

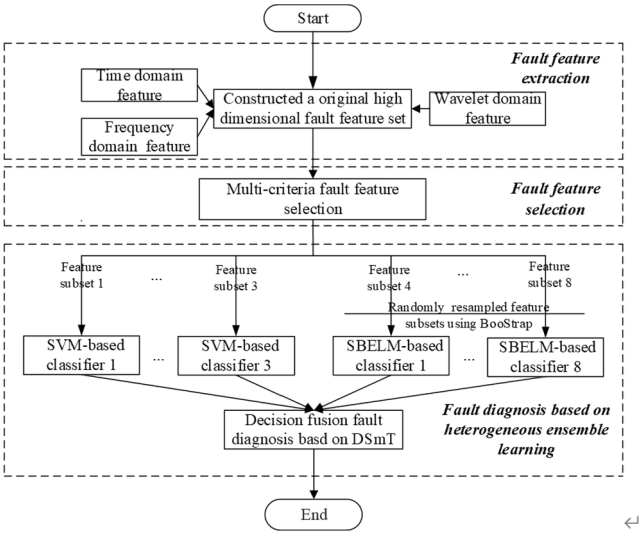


Fig. 2. Fault feature selection and decision-level information fusion

DSmT has three fundamental concepts: general frame Θ, hyper-power set D^θ, generalized belief functions.

General frame Θ: Let $\Theta = (\theta_1, \theta_2, \cdots, \theta_n)$ be a finite set of n exhaustive elements.

Hyper-power set D^θ: The Dedekinds's lattice, also named as the DSmT framework hyper power set D^θ is defined as the set of all composite propositions built from elements of Θ with ∪ and ∩ operators. In other words, $\phi, \theta_1, \theta_2, \cdots, \theta_n \in D^\theta$; If $\theta_1, \theta_2 \in D^\theta$, then $\theta_1 \cap \theta_2 \in D^\theta$ and $\theta_1 \cup \theta_2 \in D^\theta$; D^θ contains no other elements except previous.

Generalized belief functions: From a general frame Θ, give a map $m(\cdot) : D^\theta \rightarrow [0, 1]$ satisfied$m(\phi) = 0$, $m(A) > 0$ and $\sum_{A\in D^\theta} m(A) = 1$, then the quantity $m(A)$ is the generalized basic belief assignment/mas (gbba) of A. The generalized belief functions can be defined as:

$$Bel(A) = \sum_{\substack{B\subseteq A \\ B\in D^\theta}} m(B) \tag{18}$$

(b) Classic DSm rule of combination

Here the free DSm model holds for the fusion problem under consideration. For free DSm model, Θ is the exhaustive set for the problem. θ_i, $i = 1, 2, \cdots, n$ is defined as all possible assumptions for the problem. Suppose there are k independent sources of evidence over the frame Θ and their generalized belief functions are $m_1, m_2, \cdots, m_k$. Corresponds to the conjunctive consensus of the sources, for a given $A \in D^\theta$,

$$m(A) = \sum_{\substack{x_1,x_2,\cdots,x_k\in D^\theta \\ x_1\cap x_2\cap\cdots\cap x_k = A}} \prod_{i=1}^{k} m_i(x_i) \tag{19}$$

3.4. Heterogeneous ensemble learning classification

(1). Base classifier SVM use RBF kernel functions. Multiclass classification strategy is One-vs-One. Probability output is employed. Penalty parameter and RBF parameter are chosen by Particle Swarm Optimization [24].

(2). Base classifier SBELM provides different classification performance under different number of nodes in the hidden layer. The

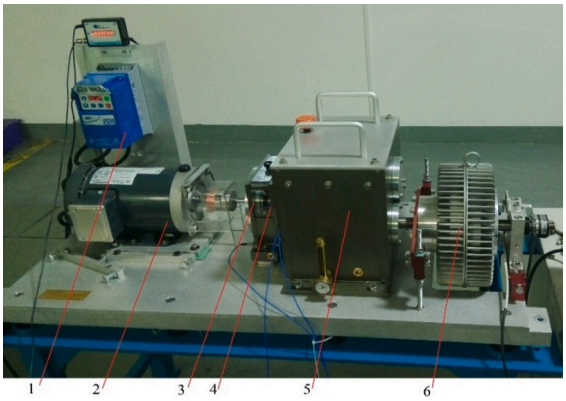


Fig. 3. Planetary gearbox experimental platform. (1-programmable control panel, 2-motor, 3- stage 1 gear train of planetary gearbox, 4- stage 2 gearbox train, 5-fixed shaft gearbox, 6-brake)

Table 1
Parameters of planetary gearbox.

Gear	Tooth number	
	Stage 1	Stage 2
Ring gear	100	100
Planetary gear (No.)	40(3)	36(4)
Sun wheel	20	28

number of nodes in the hidden layer is determined by taking 10-fold cross-validation method on each feature subset and use loop computing to calculate fault classification accuracy rate under different node numbers' condition. Node number range has been set to [10,200] with interval 10. Iterate 20 times and get a curve which shows the influence on fault classification performance from nodes changes. Then choose the optimal node number of hidden layers with highest classification accuracy rate as the result.

(3). Heterogeneous ensemble learning method firstly import 11 different quasi optimal fault feature subsets into 3 SVM and 8 SBELM. Each base classifier is trained separately with a 9:1 ratio between training samples and testing samples. Moreover, it's necessary to apply Bootstrapping random sampling strategy to obtain the training data of SBELM. The input quasi optimal feature subsets of SVM base classifiers are the subsets with top three fault classification accuracy rate.

After the training steps of 11 base classifiers, DSmT rules are applied on the probability outputs of 11 base classifiers for building a decision level fusion fault diagnosis model.

(4). Realization of fault diagnosis method

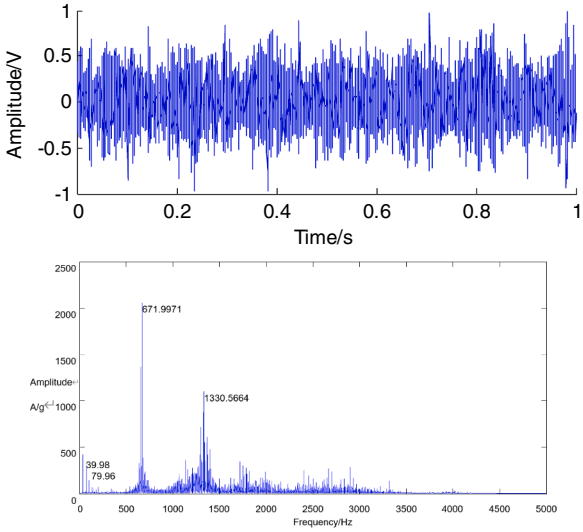
Figure 2 shows the planetary gearbox fault diagnosis process based on multi-criteria fault feature selection and heterogeneous ensemble learning classification.

4. Experiment and result discussion

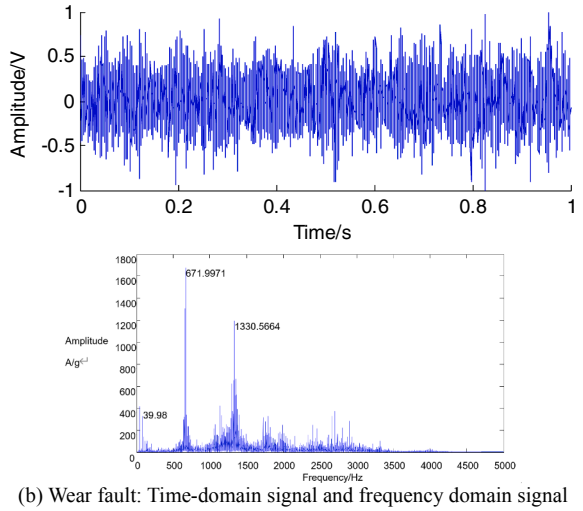
Figure 3 is the experimental platform of planetary gearbox. Planetary gearbox and fixed shaft gearbox both have two-stage gear driving structure. The characteristics of planetary gearbox is listed in Table 1. Four fault modes of sun gear are simulated in the experiment: Wear fault, Crack fault, Gnash fault and Tooth mission fault (Tooth missing fault means one of the gear teeth is completely smashed or fallen out) as indicated in Figure 4. Three accelerometer sensors are mounted externally to detect vibration signals from planetary gearbox. The sampling frequency of vibration signal is 10 kHz.

During the experiment, there are nine operating conditions of planetary gearbox which differs in rotating speed and load, include speed of 1200/2400/3600 rpm and load of 0/20/40 Nm. The continuous sampling time of each channel's vibration signal from one accelerometer sensor is 90 s under every operating condition. The single channel

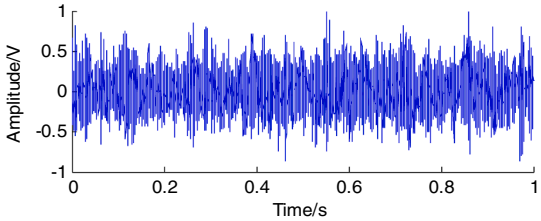
Fig. 4. Sun gear fault modes

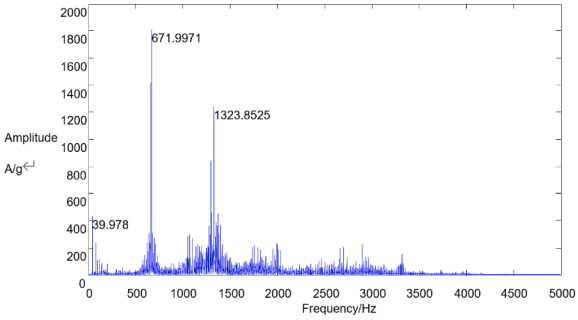


(a) Normal state: Time-domain signal and frequency domain signal

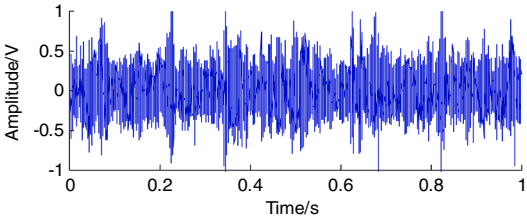


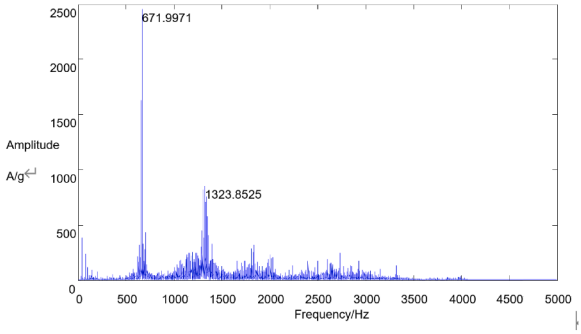
(b) Wear fault: Time-domain signal and frequency domain signal



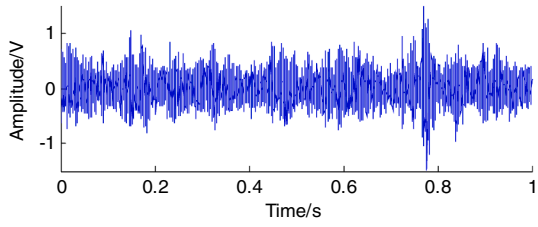


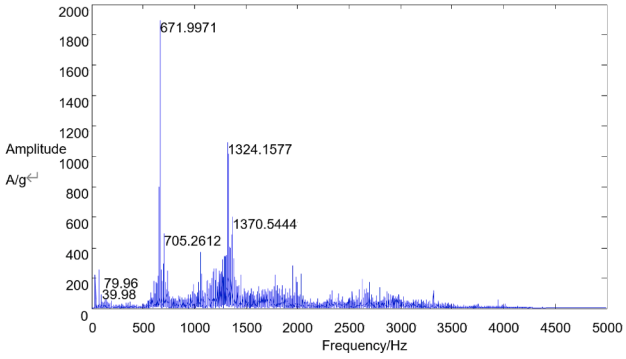
(c) Crack fault: Time-domain signal and frequency domain signal





(d) Gnash fault: Time-domain signal and frequency domain signal





(e) Tooth missing fault

Fig. 5. (a)(b)(c)(d)(e) Vibration signals in different modes of planetary gearbox

Table 2
Typical values of fault feature for different state modes.

	normal	Wear fault	Crack fault	Gnash fault	Tooth missing fault
x_1	1.010	1.065	1.044	1.428	1.486
x_2	-0.979	-1.007	-0.869	-1.404	-1.495
x_3	1.010	1.065	1.044	1.428	1.495
x_4	1.989	2.072	1.913	2.832	2.982
x_5	-1.5e-5	6.16e-05	1.86e-4	1.13e-4	-4.5e-05
x_6	0.194	0.191	0.186	0.199	0.191
x_7	0.163	0.159	0.157	0.168	0.159
x_8	0.060	0.059	0.055	0.064	0.061
x_9	0.264	0.218	0.200	0.221	0.287
x_{10}	0.245	0.243	0.234	0.254	0.247
x_{11}	3.180	3.375	3.188	3.623	4.150
x_{12}	0.082	0.101	0.163	0.128	0.217
x_{13}	1.264	1.277	1.260	1.274	1.294
x_{14}	4.112	4.369	4.453	5.620	6.049
x_{15}	5.198	5.579	5.611	7.160	7.831
x_{16}	6.165	6.666	6.639	8.509	9.378
x_{17}	16.74	17.92	18.99	22.11	24.47
x_{18}	12.11	12.89	11.59	12.31	13.16
x_{19}	1700	1698	1695	1690	1735
x_{20}	1918	1925	1931	1917	1973
x_{21}	889	906	925	905	939
x_{22}	0.112	0.116	0.131	0.145	0.080
x_{23}	0.419	0.367	0.421	0.430	0.378
x_{24}	0.066	0.0811	0.0579	0.059	0.065
x_{25}	0.336	0.356	0.330	0.309	0.409
x_{26}	1.50e-4	1.47e-4	2.29e-4	1.62e-4	2.68e-4
x_{27}	0.002	0.003	0.0026	0.002	0.003
x_{28}	0.052	0.063	0.042	0.043	0.046
x_{29}	0.011	0.013	0.0141	0.011	0.018
x_{30}	1.374	1.436	1.372	1.374	1.349
x_{31}	149	153	159	178	126
x_{32}	220	215	236	201	199
x_{33}	108	115	92	101	104
x_{34}	202	217	203	183	182
x_{35}	0.712	0.773	0.855	0.767	1.154
x_{36}	7.353	8.916	8.241	7.820	11.26
x_{37}	89.82	98.89	73.84	83.12	84.24
x_{38}	27.67	31.90	32.77	30.65	40.20
x_{39}	1.734	1.760	1.744	1.733	1.754

signal's sampling data are divided into 3865 data segments while the length of each segment is 10000 points. In the end, extract 39 fault feature parameters from each data segment and form a 3865-sets×39-dimension fault feature sampling space, with the first 3480 fault feature sets taken as training samples while last 385 chosen as test samples. Software is MATLAB 2013a, i7-6700 operating computer has 8 cores CPU with 3.4 GHz frequency and 16 GB Memory.

Figure 5 shows the time-domain and frequency-domain waveform of the first vibration signal operating under condition 2400 rpm, 20 Nm. 5 typical values of fault feature corresponding to 5 state modes are listed in Table 2.

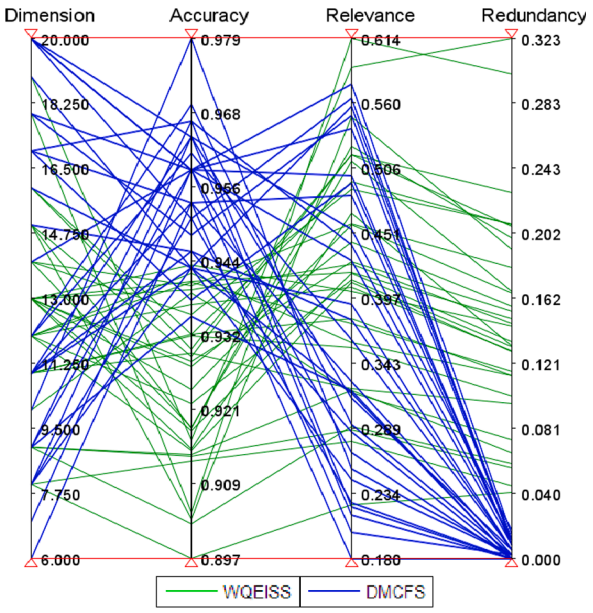


Fig. 6. Parallel coordinates of feature subsets selected by WQEISS (Green line, NSGA-II) and DMCFS (Blue line, MOEA/D) considering 4 target function (Dimension, Accuracy, Relevance, Redundancy)

4.1. Fault feature selection results

To fully test the effectiveness and advantages of the Diverse Multi-Criteria Feature Selection (DMCFS) algorithm proposed in this paper, following fault feature selection algorithms are chosen for comparison:

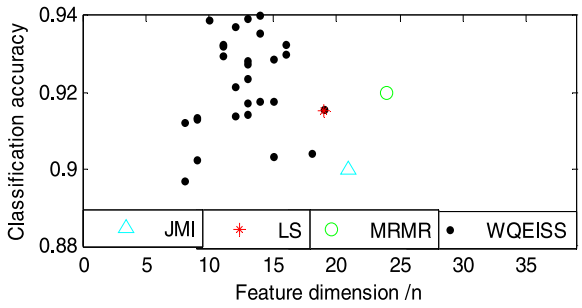
)1. Feature selection based on Joint Mutual Information, JMI [25]. Fault feature selection criterion is JMI, the mathematic model for selecting optimal feature subset is built up.
)2. Feature selection method based on Laplacian Score, LS [26]. Fault feature selection criterion is Laplacian Score, then build up the mathematic model for selecting optimal feature subset.
)3. Feature selection based on Maximum Relevance and Minimum Redundancy, MRMR [27]. Fault feature selection criterion is relevance and redundancy, the mathematic model for selecting optimal feature subset is:

$$\begin{cases} minF(S) \\ F(S) = \{-f_1(S), f_2(S)\} \end{cases} \tag{18}$$

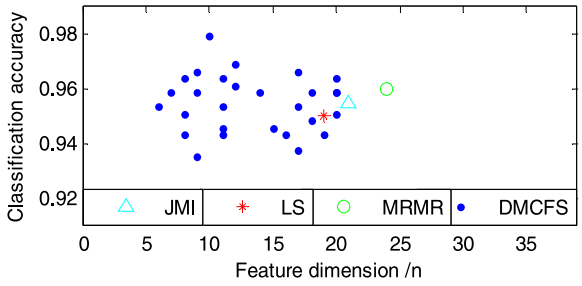
4) Wrapper for Quasi Equally Informative Subset Selection, W-QEISS [28]. W-QEISS is universal feature selection algorithm without considering specific object. It takes ELM to evaluate classification

Table 3
Feature selection results of 5 fault feature selection algorithms.

Algorithm	Feature serial number/Subset vector dimension
JMI	[5;34;32;30;39;23;11;12;13;19;20;24;21;26;27;28;29;14;22;16]/21
LS	[4;18;36;8;9;10;6;38;7;3;2;35;1;17;20;19;27;33;37]/19
MRMR	[33;23;5;12;21;14;34;28;31;1;13;37;39;35;15;27;32;30;38;11;3;16;24;25]/24
WQEISS	[6;7;9;18;20;21;22;28;30;36;37;38;39]/13
	[6;7;9;10;18;20;21;22;29;30;36;37;38;39]/14
	[6;7;8;9;10;18;20;21;22;29;36;37;38;39]/14
	[6;9;10;18;20;21;22;29;30;36;37;38;39]/13
	[6;7;9;18;20;21;22;29;30;36;37;38]/12
DMCFS	[7;17;18;24;27;31;32;37;38;39]/10
	[6;9;18;19;20;23;26;32;33 37 38;39]/12
	[9;10;17;18;32;38;39]/7
	[7;9;10;17;18;24;26;27;32;33;39]/11
	[7;10;17;18;32;39]/6



(a) Feature classification results by W-QEISS, in comparison with JMI, LS, MRMR



(b) Feature classification results by DMCFS, in comparison with other methods

Fig. 7. Performance analysis and comparison of proposed quasi optimal feature subsets extraction method

performance and acquires multiple feature subsets through Non-Dominated Soring Genetic Algorithm II (NDSGA) with elitism.

Fault feature selection results obtained from different feature selection algorithms are shown in Table 3.

In Table 3, the serial number of selected feature parameter are the subscripts of original 39-dimensional fault features $x_1 \sim x_{39}$. With regard to feature selection results by using W-QEISS and DMCFS algorithm, only the top 5 fault feature subsets with highest classification accuracy rate are listed. It can be obtained that the fault feature subsets acquired by JMI, LS and MRMR have high dimensions and remains some redundant fault features. Combined with Figure 6, difference between the feature subsets selected by W-QEISS algorithm is very small. However, their dimension and redundancy of feature subsets are high, and its feature diversity is obviously insufficient. Moreover, its classification accuracy has negligible differences against JMI, LS and MRMR. The main reason is that the NSGA-II method employed by W-QEISS has disadvantages of low evolutionary efficiency, slow convergence and uneven individual distribution. Another reason which leads to poor classification performance is that it applies ELM classifier to evaluate classification performance. On the other hand, the feature subsets obtained by DMCFS algorithm dominates in feature dimension, classification accuracy rate and redundancy compared with W-QEISS. It also has significant superiority in selected feature subsets which leads to a better diversity.

The fault classification accuracy of optimal feature subsets obtained by JMI, LS, MRMR, W-QEISS and DMCFS are listed in Figure 7. Comparison is shown in Table 4.

Table 4
Characteristics comparison between 5 algorithms for optimal feature selection.

Algorithm	Best classification accuracy (%)	Feature dimension	Number of hidden layer nodes	Run time (s)
JMI	90.01	21	100	0.5
LS	91.12	19	100	60
MRMR	92.00	24	100	0.5
WQEISS	94.05	13	100	1574
DMCFS	97.91	10	60	130

Table 5
Diagnosis results of different fault diagnosis algorithms under selected optimal feature subsets.

Algorithm	Classification accuracy	Algorithm	Classification accuracy
JMI+ELM	0.9001	WQEISS+EELM	0.9557
JMI+SVM	0.9545	WQEISS+ESVM	0.9601
JMI+EELM	0.9241	DMCFS+SVM	0.9613
MRMR+ELM	0.9200	DMCFS+SBELM	0.9791
MRMR+SVM	0.9577	DMCFS+EELM	0.9574
MRMR+SBELM	0.9600	DMCFS+ESVM	0.9629
MRMR+EELM	0.9466	Original features +EELM	0.9274
WQEISS+ELM	0.9405	Original features +ESVM	0.9462
WQEISS+SVM	0.9682		

Figure 6(a) uses ELM classifier to evaluate the classification accuracy. Figure 6(b) uses SBELM classifier to evaluate the classification accuracy.

In Table 4, ELM classifier are used to evaluate the classification accuracy of selected feature subsets by JMI, LS, MRMR.

By analyzing Figure 7 and Table 4, following results are summarized:

1) Multi-criteria feature selection method helps acquire lower dimensional fault feature subsets and gets better fault classification performance.
2) Multi-objective evolutionary algorithm leads to diverse multi-criteria fault feature subsets. By analyzing the similarity of feature subsets selected by different algorithm: $x_{\widehat{7}}$, $x_{\widehat{9}}$, $x_{\widehat{10}}$, $x_{\widehat{20}}$, $x_{\widehat{23}}$, $x_{\widehat{30}}\cdots$ are strong correlation features, $x_{\widehat{1}} \sim x_{\widehat{5}}$, $x_{\widehat{12}}$, $x_{\widehat{13}}$ are redundant features. Redundant features exist in approximately 7-12 dimensions.
3) Compared with W-QEISS, DMCFS dominates in optimization speed and calculation efficiency. DMCFS obtains multiple quasi optimal feature subsets with better diversity. These feature subsets meet the requirement of fault classification accuracy. They have low redundancy, less dimensions and high relevancy.

Table 6
Fault diagnosis results of different fault diagnosis algorithms under diverse quasi optimal feature subsets.

Algorithm	Classification accuracy	Algorithm	Classification accuracy
JMI+HELC	0.9774	WQEISS+HELC	0.9800
MRMR+HELC	0.9711	DMCFS+EELM	0.9817
WQEISS+EELM	0.9630	DMCFS+ESVM	0.9752
WQEISS+ESVM	0.9514	DMCFS+HELC	0.9962

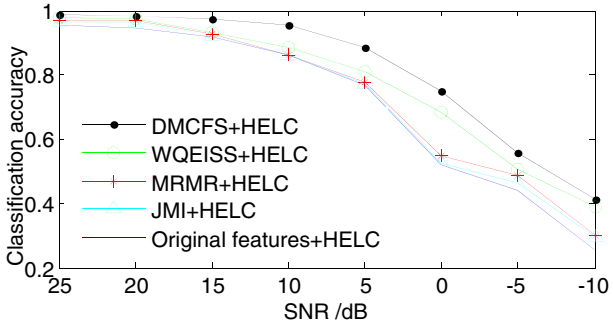


Fig. 8. Fault diagnosis results of HELC algorithm under different signal-to-noise ratio

4.2. Fault diagnosis results

This part verifies the effectiveness and advantage of Heterogeneous Ensemble Learning Classification (HELC) method developed in this paper through 3 aspects.

1) Fault diagnosis performance with multi-criteria optimal feature subsets

Separately input 3 optimal feature subsets obtained by comparing fault feature selection algorithm and DMCFS into base classifier and ensemble learning classifier to estimate their different fault diagnosis performance. Results are listed in Table 5.

In Table 5, the optimal feature subset obtained by WQEISS and DMCFS are used for the input to fault classifiers. Ensemble Extreme Learning Machine (EELM) takes Bagging strategy and ELM base classifier. Ensemble Support Vector Machine (ESVM) is based on Bagging strategy and SVM base classifier.

Thus, the stability of fault diagnosis result (variation range of classification accuracy) of ELM and SBELM is worse compared with SVM. The change of fault diagnosis performance under different feature subsets by SVM is very small while ELM is large and SBELM is the medium one.

Applying simple Bagging strategy on SVM rarely improve its diagnosis performance, with the classification accuracy always around 96%. However, Bagging strategy does better improvement on EELM. Meanwhile, using single optimal feature subset to form ensemble learning model only has little promotion on fault diagnosis accuracy rate.

2) Fault diagnosis performance under diverse multi-criteria quasi optimal feature subsets

Here takes the optimal feature subset acquired by JMI and MRMR, the diverse quasi optimal feature subsets acquired by W-QEISS and DMCFS as input samples into EELM, ESVM and HELC classifiers, in order to determine their joint fault diagnosis performance with diverse quasi optimal feature subsets and heterogeneous ensemble learning classifier. Results are shown in Table 6.

In Table 6, the diverse quasi optimal fault feature subsets obtained by WQEISS and DMCFS are the input of both fault classifiers. Thus, the fusion method which combines diversified quasi optimal fault feature subsets and heterogeneous ensemble learning classifier can enhance fault diagnosis accuracy. The maximum diagnosis accuracy reaches 99.62% which verifies the feasibility of the algorithm mentioned above.

3) Fault diagnosis performance under different noise conditions

In real operating condition, the vibration signals of planetary gearbox will be affected by internal or external environmental noise, for example they often operate under the working condition that colored noise disturbance frequently exists. To simulate real environment, add colored noise to sampling data of original vibration signal to form signals with noise. The colored noise is generated by international common noise library NoiseX-92 and chosen to be engine room noise [29]. MRMR, W-QEISS and DMCFS algorithms are taken for optimal feature selection. Then carry on fault diagnosis based on HELC to determine this algorithm's robustness against noises. Results shown in Figure 7.

From Figure 8, along with the decrease in Signal-to-Noise Ratio (SNR), fault diagnosis performance of HELC algorithm with different optimal feature subsets obtained by DMCFS, WQEISS, MRMR, JMI and original high dimension feature set keeps going worse. However, the method using DMCFS feature selection method and HELC classifier has highest fault diagnosis accuracy in the range 10-25 dB of SNR. So the proposed fault diagnosis method has a strong robustness.

5. Conclusions

This paper proposed an effective and powerful method for precise fault diagnosis of planetary gearbox, which based on diverse multi-criteria feature selection and heterogeneous ensemble learning classification. The criteria for selecting appropriate fault features from multi-domain high dimensional fault features of planetary gearbox is designed, and the mathematic model of multi-criteria fault feature selection is established. The diverse low-dimensional quasi optimal fault feature subsets are obtained by MOEA/D. Every quasi optimal fault feature subset is transferred to a base classifier for primary fault diagnosis. The outputs of the base classification are used as the inputs to Dezert-Smarandache rules for classifier-level information fusion. The proposed method for fault diagnosis of planetary gearbox is proven to be effective and accurate through several experiments. The achieved results state clearly that the proposed method has following advantages: (1) Well-designed fault feature selection criteria help find multiple optimal fault feature subsets which are low-dimensional and irredundant; (2) Diverse quasi optimal fault feature subsets can be obtained by MOEA/D; (3) The Dezert-Smarandache theory of evidence demonstrates its capability to incorporate the uncertain output data of base classifier into precise and reliable fault classification application; (4) The decision-level information fusion based fault diagnosis method which employs diverse quasi optimal fault feature subsets, heterogeneous ensemble learning classifier and Dezert-Smarandache theory of evidence truly it highly improve the overall fault diagnosis accuracy for planetary gearbox.

In summary, the feature selection method based on MOEA/D evolutionary algorithm can be used to reduce the fault feature size and obtain multiple quasi optimal feature subsets. By integrating feature layer fusion and classifier-level fusion, the proposed method can effectively diagnose early faults in planetary gearbox and improve fault diagnosis performance under noise condition.

CRediT authorship contribution statement

Zirui Wang: Conceptualization, Methodology, Validation, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Haian Huang:** Data curation, Software, Formal analysis. **Youren Wang:** Resources, Supervision, Project administration.

Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work. There is no professional or other personal interest of any nature or kind in any product, service and/or company that could be constructed as influencing the position presented in, or the review of, the manuscript entitled.

Acknowledgement

The authors would like to thank the editor/ associate editor and the anonymous reviewers for their insightful comments.

This work is supported in part Foundation of National Key Laboratory of Science and Technology on Helicopter Transmission under project (No. KY-52-2018-0024), the Aviation Science Foundation of China (No. 20183352031), and Innovation Foundation of COMAC Engineering and Technological Research Centre (No. SAMC14-JS-15-051).

Appendix

(See Table A1)

Table A1
Multi domain fault feature characteristics.

Feature	Name	Definition
x_1	maximal value	Maximal value of a time domain signal,x_{max}
x_2	minimum value	Minimum value of time domain signal,x_{min}
x_3	peak value	$= max\{\|x_i\|\}$
x_4	peak to peak value	$x_{max} - x_{min}$
x_5	mean value	$\frac{1}{N}\sum_{i=1}^{N} x_i$, N is signal length
x_6	average absolute value	$\frac{1}{N}\sum_{i=1}^{N} \|x_i\|$
x_7	Square mean root	$= \left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\|x_i\|}\right)^2$
x_8	variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$
x_9	standard deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$
x_{10}	root mean square	$= \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
x_{11}	kurtosis	$\frac{1}{N\sigma^4}\sum_{i=1}^{N}(x_i - \overline{x})^4$ σ is standard deviation
x_{12}	skewness	$\frac{1}{N\sigma^3}\sum_{i=1}^{N}(x_i - \overline{x})^3$
x_{13}	shape factor	$= \frac{x_{rms}}{\frac{1}{N}\sum_{i=1}^{N}\|x_i\|}$
x_{14}	crest factor	$= \frac{x_{peak}}{x_{rms}}$
x_{15}	impulse factor	$= \frac{x_{peak}}{\frac{1}{N}\sum_{i=1}^{N}\|x_i\|}$
x_{16}	margin factor	$= \frac{x_{peak}}{\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\|x_i\|}\right)^2}$
x_{17}	clearance factor	$= \frac{x_{peak}}{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
x_{18}	mean frequency	$\frac{1}{N}\sum_{i=1}^{N} A_i$, A_i is Y-axis value after FFT
x_{19}	central frequency	$\frac{\sum_{i=1}^{N}(f_i * A_i)}{\sum_{i=1}^{N} A_i} f_i$is X-axis value after FFT
x_{20}	frequency RMS value	$\sqrt{\frac{\sum_{i=1}^{N}(f_i^2 * A_i)}{\sum_{i=1}^{N} A_i}}$
x_{21}	frequency standard deviation	$\sqrt{\frac{\sum_{i=1}^{N}((f_i - x_{19})^2 * A_i)}{\sum_{i=1}^{N} A_i}}$
Wavelet domain	x_{22}~x_{39} are the features after db6 wavelet transform, with the decomposition level set to 3, which leads to 8 sub bands.	Assume each sub band energy is $E_3^j, j = 1, 2, 3, \cdots, 8$, then the energy equation$E_3^j = \sum_{k=1}^{N}\|u_{jk}\|^2 u_{jk} (j = 1, 2, \cdots, 8; k = 1, 2, \cdots, N)$ represents the third sub band j_{th}node k_{th}discrete point value.Total energy$E = \sum_{i=1}^{8} E_3^i$
x_{22}~x_{29}	wavelet energy ratio	$x_{22} = \frac{E_3^1}{E}$ $x_{23} = \frac{E_3^2}{E}$ $x_{24} = \frac{E_3^3}{E}$ $x_{25} = \frac{E_3^4}{E}$ $x_{26} = \frac{E_3^5}{E}$ $x_{27} = \frac{E_3^6}{E}$ $x_{28} = \frac{E_3^7}{E}$ $x_{29} = \frac{E_3^8}{E}$
x_{30}	wavelet energy entropy	$-\sum_{j=1}^{8}\left(\frac{E_3^j}{E} log\left(\frac{E_3^j}{E}\right)\right)$

(continued on next page)

Table A1 (*continued*)

Feature	Name	Definition
x_{31}~x_{38}	wavelet feature scale entropy	$x_{31} = -\sum_{k=1}^{N} \lvert u_{1k}\rvert^2 log(\lvert u_{1k}\rvert^2)$ $x_{32} = -\sum_{k=1}^{N} \lvert u_{2k}\rvert^2 log(\lvert u_{2k}\rvert^2)$ $x_{33} = -\sum_{k=1}^{N} \lvert u_{3k}\rvert^2 log(\lvert u_{3k}\rvert^2)$ $x_{34} = -\sum_{k=1}^{N} \lvert u_{4k}\rvert^2 log(\lvert u_{4k}\rvert^2)$ $x_{35} = -\sum_{k=1}^{N} \lvert u_{5k}\rvert^2 log(\lvert u_{5k}\rvert^2)$ $x_{36} = -\sum_{k=1}^{N} \lvert u_{6k}\rvert^2 log(\lvert u_{6k}\rvert^2)$ $x_{37} = -\sum_{k=1}^{N} \lvert u_{7k}\rvert^2 log(\lvert u_{7k}\rvert^2)$ $x_{38} = -\sum_{k=1}^{N} \lvert u_{8k}\rvert^2 log(\lvert u_{8k}\rvert^2)$
x_{39}	wavelet singular entropyLet J=8. $A_{J\times N}$ is coefficient matrix, N is sample number. Apply SVD:$A_{J\times N} = U_{J\times K} W_{K\times K} V_{K\times N} W_{K\times K}$ is diagonal matrix.$\lambda_i (i = 1, 2, \cdots K^{'}), K^{'} \le K, K^{'} \le J$ is the SV of $A_{J\times N}$.	$x_{39} = -\sum_{i=1}^{K^{'}} p_i log(p_i) p_i = \frac{\lambda_i}{\sum_{j=1}^{K^{'}} \lambda_j}$

References

[1] Z. Feng, L. Zhao, F. Chu, Vibration spectral characteristics of localized gear fault of planetary gearboxes[J], Proceedings of the CSEE 33 (5) (2013) 119–127.

[2] Y. Lei, J. Lin, M.J. Zuo, et al., Condition monitoring and fault diagnosis of planetary gearboxes: A review[J], Measurement 48 (1) (2014) 292–305.

[3] X. Liang, M.J. Zuo, L. Liu, A windowing and mapping strategy for gear tooth fault detection of a planetary gearbox[J], Mechanical Systems & Signal Processing 80 (2016) 445–459.

[4] Y. Lei, Z. Liu, X. Wu, et al., Health condition identification of multi-stage planetary gearboxes using a mRVM-based method[J], Mechanical Systems & Signal Processing 60–61 (2015) 289–300.

[5] Y. Li, Y. Yang, G. Li, et al., A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection[J], Mechanical Systems & Signal Processing 91 (2017) 295–312.

[6] F. Chen, B. Tang, Z. Shu, Rotating machinery fault diagnosis based on isometric mapping and weighted KNN[J], Chinese Journal of Scientific Instrument 34 (1) (2013) 215–220.

[7] T. Wang, Q. Han, F. Chu, et al., Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review[J], Mechanical Systems and Signal Processing 126(JUL.1):662–685 (2019).

[8] M. Cerrada, G. Zurita, D. Cabrera, et al., Fault diagnosis in spur gears based on genetic algorithm and random forest[J], Mechanical Systems & Signal Processing 70–71 (2015) 87–103.

[9] Z. Liu, M.J. Zuo, H. Xu, Fault diagnosis for planetary gearboxes using multi-criterion fusion feature selection framework, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 227 (9) (2013) 2064–2076.

[10] B. Tang, J. Ma, Manifold learning method for fault diagnosis based on sensitive feature selection with multi-criteria evaluation sequences and adaptive neighborhood[J], Chinese Journal of Scientific Instrument 8 (4) (2014) 377–401.

[11] F. Pacheco, M. Cerrada, R.V. Sánchez, et al., Attribute clustering using rough set theory for feature selection in fault severity classification of rotating machinery[J], Expert Systems with Applications 71 (2017) 69–86.

[12] C. Li, R.V. Sanchez, G. Zurita, et al., Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals[J], Mechanical Systems & Signal Processing 76 (2016) 283–293.

[13] X. Wang, R. Zhao, L. Deng, Rotating machinery fault diagnosis based on KSLPP and RWKNN[J], Journal of Vibration and Shock 35 (8) (2016) 219–223.

[14] J. Yu, Machinery fault diagnosis using joint global and local/nonlocal discriminant analysis with selective ensemble learning[J], Journal of Sound & Vibration 382 (2016) 340–356.

[15] Q. Zhang, H. Li, MOEA/D: A Multi-objective Evolutionary Algorithm Based on Decomposition[J], IEEE Transactions on Evolutionary Computation 11 (6) (2007) 712–731.

[16] M. Khazaee, H. Ahmadi, M. Omid, et al., Classifier fusion of vibration and acoustic signals for fault diagnosis and classification of planetary gears based on Dempster-Shafer evidence theory[J], Journal of Process Mechanical Engineering 228 (1) (2014) 21–32.

[17] Z. Liu, Q.u. Jian, M.J. Zuo, X.u. Hong-bing, Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and kernel Fisher discriminant analysis[J], The International Journal of Advanced Manufacturing Technology 67 (5) (2013) 1217–1230.

[18] L. Zhao, X. Liu, L. Lou, The Feature extraction method of non-stationary vibration signal based on SVD-Complex analytical wavelet demodulation[J], Journal of Vibration, Measurement & Diagnosis 35 (04) (2015) 672–676.

[19] F. Feng, Early Fault Diagnosis Technology for Bearing Based on Wavelet Correlation Permutation Entropy[J], Journal of mechanical engineering 48 (13) (2012) 73–79.

[20] J. Luo, C.M. Vong, P.K. Wong, Sparse Bayesian extreme learning machine for multi-classification[J], IEEE Transactions on Neural Networks & Learning Systems 25 (4) (2014) 836–843.

[21] E. Soriaolivas, J. Gomezsanchis, J.D. Martin, et al., BELM: Bayesian Extreme Learning Machine[J], IEEE Transactions on Neural Networks 22 (3) (2011) 505–509.

[22] F.R. Burden, M.G. Ford, D.C. Whitley, et al., Use of Automatic Relevance Determination in QSAR Studies Using Bayesian Neural Networks[J], Journal of Chemical Information & Computer Sciences 40 (6) (2013) 1423–1430.

[23] X.X. Xiangjun Li, YU.X. Zhuang, et al., Research on a Method of Multiple Faults Diagnosis Based on Hierarchical DSmT[J], Control and Decision 5 (2016) 875–881.

[24] Libsvm-Faruto, Toolbox address: http://www.ilovematlab.cn/thread-63767-1-1.html.

[25] M. Bennasar, Y. Hicks, R. Setchi, Feature selection using Joint Mutual Information Maximisation[J], Expert Systems with Applications 42 (22) (2015) 8520–8532.

[26] X. He, D. Cai, P. Niyogi, Laplacian Score for Feature Selection[J], Advances in Neural Information Processing Systems 18 (2005) 507–514.

[27] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: Criteria of max-dependency, max-relevance and min-redundancy[J], IEEE Transactions on Pattern Analysis and Machine Intelligence 2 (8) (2005) 1226–1238.

[28] G. Karakaya, S. Galelli, S.D. Ahipasaoglu, et al., Identifying (Quasi) Equally Informative Subsets in Feature Selection Problems for Classification: A Max-Relevance Min-Redundancy Approach[J]. IEEE Transactions on, Cybernetics (2016:1.).

[29] P.D. Swami, R. Sharma, A. Jain, et al., Speech enhancement by noise driven adaptation of perceptual scales and thresholds of continuous wavelet transform coefficients[J], Speech Communication 70 (2015), 1 1.