Mechanical Systems and Signal Processing 28 (2012) 608–621

Contents lists available at SciVerse ScienceDirect

Mechanical Systems and Signal Processing

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

Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine

Ning Li [a,*], Rui Zhou [b], Qinghua Hu [c], Xiaohang Liu [a]

[a] Shanghai Second Polytechnic University, Shanghai 201209, PR China
[b] China Ship Development and Design Center, Shanghai 201108, PR China
[c] Harbin Institute of Technology, Harbin, Heilongjiang 150001, PR China

ARTICLE INFO

Article history:
Received 2 January 2011
Received in revised form
7 October 2011
Accepted 14 October 2011
Available online 15 November 2011

Keywords:
Redundant second generation wavelet packet transform
Neighborhood rough set
Support vector machine
Attribute reduction
Fault diagnosis

ABSTRACT

This paper investigates the application of the redundant second generation wavelet package transform (RSGWPT), neighborhood rough set (NRS) and support vector machine (SVM) on faulty detection, attribute reduction and pattern classification. On this basis, a novel method for mechanical faulty diagnosis based on RSGWPT, NRS and SVM is presented, which utilizes the RSGWPT to extract faulty feature parameters from the statistical characteristics of wavelet package coefficients to constitute feature vectors, and then makes the attribute reduction by NRS method to obtain the key features, lastly these key features are input into SVM to accomplish faulty pattern classification. The experimental results of the proposed method to fault diagnosis of the gearbox and gasoline engine valve trains show that this method can extract the faulty features, which have better classification ability and at the same time reduce a lot of redundant features in case of assuring the classification accuracy, accordingly improve the classifier efficiency and achieve a better classification performance.

1. Introduction

With the growing demand of high quality production, the effective mechanical faulty diagnosis has been gaining increasing attention [1–3]. Generally, there are two important steps in the faulty diagnosis system: the first is the signal processing for the faulty feature extraction, which is to determine wheather a fault has occurred in the equipment, and the second step consists of the pattern classification based on the characteristics obtained in the previous step. Presently when evaluating the effectiveness of a fault diagnosis method, the accuracy and speed are both very important factors, so finding an accurate and fast method for fault diagnosis is an essential issue [4,5]. In a word, the purpose of mechanical faulty diagnosis research is to find a method that can accurately evaluate the faulty type using the least faulty features extracted by special signal processing method from the collected vibration signal or sound emission signal.

The incipient faulty feature is often weak and buried in the background signal, so it is difficult for the traditional signal processing method to detect them. Second generation wavelet transform (SGWT), proposed by Wim Sweldens, is a new wavelet construction method using lifting scheme [6,7]. It can be seen as an alternate implementation of classical discrete wavelet transform. The main feature of SGWT is that it provides an entirely spatial domain interpretation of the transform,

* Corresponding author. Tel./fax: +86 21 50217018.
E-mail address: ningligood@gmail.com (N. Li).

0888-3270/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ymssp.2011.10.016

as opposed to the traditional frequency domain based constructions. The time-frequency resolution of SGWT varies with the decomposition levels. It gives good time and poor frequency resolution at high-frequency subband, and good frequency and poor time resolution at low-frequency subband. In order to obtain a higher resolution in the high-frequency subband which the faulty characteristics always exist in, second generation wavelet package transform (SGWPT) has been constructed and hence the detail coefficients at each level are further decomposed to obtain their approximation and detail components [4,8–10]. Unfortunately, SGWPT do not have time invariant property. The decomposition results of a delayed signal are not the time-shifted version of those of the original signal. This may result in the loss of useful faulty information for feature extraction and fault diagnosis. The redundant lifting scheme possesses time invariant property and overcomes the disadvantage of lifting scheme by getting rid of the split step and zero padding of prediction operator and update operator, which makes the approximation and detail signals at all levels are the same length as the original signal [11–13]. The redundant lifting scheme based wavelet packet transform can not only afford more detailed local time-frequency description of the signal, but also inhibit the frequency aliasing components of the analysis result because of the absence of the split and merge step in the decomposition and reconstruction stage [14]. Consequently, the statistical features extracted from the transform coefficients of RSGWPT have a greater ability to detect the faulty signal.

Generally the vibration signals are acquired from the monitoring mechanical equipments. After A/D conversion, the sampled vibration data decomposed using RSGWPT. From each of the resultant sub-band wavelet packet coefficients, the statistical features can be calculated to describe the characteristics of the signal at each scale. These statistical features can be applied in the fault diagnosis directly, but in practice some of the features extracted from experimental data are usually imperfect and redundancy, even incompatible for each other, accordingly bring forth a lot of problems to mechanical fault diagnosis such as high computational complexity, low recognition speed and weak recognition effect. So it is meaningful to investigate the fault diagnosis method using the less attribute values without missing any fault information.

Rough set theory is a useful mathematical tool to solve the problem of uncertain, imprecise and vague information, which regards the knowledge is about the domain compartition and has the property of granularity. The characteristics of rough set theory are creating approximate descriptions of objects for data analysis, optimization and recognition, and it does not need the domain or prior knowledge. Therefore using this method can evaluate the importance of various attributes and retain some key attributes with no additional knowledge except for the supplied data required [5]. Recently, the rough sets approach has been applied in many domains, such as machine fault diagnosis, stock market forecast, decision support systems, medical diagnosis, data filtration and software engineering [15–18].

However, the classical rough set model can just be used to process categorical features that are discrete values, so for the rough set based feature selection in mechanical fault diagnosis system, it is necessary to introduce a discretizing algorithm to partition the value domains of real-valued variables into several intervals and then regard them as categorical features. Many discretization methods of numerical attributes have been proposed in recent years, including equidistance method, equal frequency method, maximum entropy method, etc. [16]. Obviously, discretization of numerical attributes

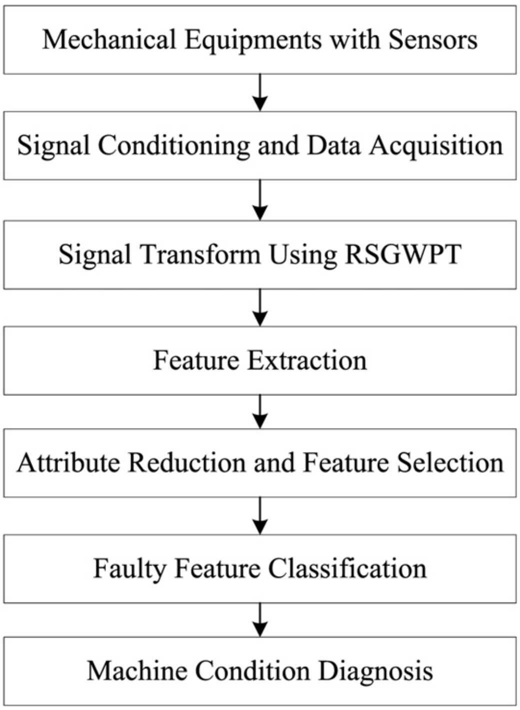

Fig. 1. Flow diagram of the proposed fault diagnostic procedure.

may cause information loss because the degrees of membership of numerical values to discretized values are not considered [19,20]. For solving this problem, a neighborhood rough set model, which can deal with both categorical and numerical attributes, was introduced and some attribute reduction algorithms based on this model have been constructed [20–22]. The neighborhood rough set model based attribute reduction can perfectly preserve the fault information of original data because of the absence of discretizing numerical attributes.

For classification, there is a crowded literature about the selection of the optimal classifiers for specific applications such as gear fault detection, reciprocating compressor, roller bearing, etc. [23–26]. Support vector machine (SVM) is a novel machine learning method based on statistical learning theory originally introduced by Vapnik, which has many attractive features including its desirable classification ability with a small quantity of fault data samples and promising empirical performance in the field of non-linear and high-dimensional pattern recognition [27–29].

According to the above principles, a new fault diagnosis method for mechanical equipment is proposed in this paper, which utilizes the RSGWPT constructed based on redundant lifting scheme and SGWPT to extract faulty feature parameters and then makes the attribute reduction in the faulty feature vectors by NRS method to obtain the key features, lastly these key features are input into SVM to accomplish faulty pattern classification. The flow diagram of the proposed fault diagnosis method is shown in Fig. 1.

The rest of this paper is organized as follows: in Section 2, the statistics parameters of the wavelet package transform coefficients are extracted as faulty features by RSGWPT. In Section 3, the feature vectors are reduced by NRS and some key features are selected to be classified. In Section 4, we use SVM to accomplish the fault pattern classification based on the selected features. In Sections 5 and 6, the proposed method is applied to diagnose different states of a gearbox and the valve train on a gasoline engine, and the results have verified the validity of the proposed diagnosis method. At last the conclusions have been briefly drawn in Section 7.

2. RSGWPT and fault feature extraction

For the purpose of implementing the construction of RSGWPT, SGWPT and the redundant lifting scheme are introduced in this section.

Firstly the decomposition and reconstruction stages of SGWPT are described as below. The decomposition stage of SGWT consists of three steps: split, prediction and update. In the decomposition stage, $X_{l,k}$ is split into even samples $X_{l,ke}$ and odd samples $X_{l,ko}$,

$$X_{l,ke} = X_{l,k}(2i), \quad X_{l,ko} = X_{l,k}(2i+1) \tag{1}$$

where $X_{l,k}$ represents the coefficients of the kth node at level l.

Then calculate each subband coefficients at level $l+1$ according to the following formula:

$$\begin{cases} X_{l+1,2} = X_{l,1o} - P(X_{l,1e}) \\ X_{l+1,1} = X_{l,1e} + U(X_{l+1,2}) \\ \cdots \\ X_{l+1,2^{l+1}} = X_{l,2^l o} - P(X_{l,2^l e}) \\ X_{l+1,2^{l+1}-1} = X_{l,2^l e} + U(X_{l+1,2^{l+1}}) \end{cases} \tag{2}$$

where P is the designed prediction operator and applied on $X_{l,2^l e}$ to predict $X_{l+1,2^{l+1}}$. The resultant prediction error $X_{l+1,2^{l+1}}$ is regarded as the wavelet package coefficients of the 2^{l+1}th node at level $l+1$. U is the designed update operator and applied on $X_{l+1,2^{l+1}}$. Adding the result to the even samples, the resultant $X_{l+1,2^{l+1}-1}$ is regarded as wavelet package coefficients of the $(2^{l+1}-1)$th node at level $l+1$. The prediction operator P and update operator U are built by means of interpolating subdivision method (ISM) [30,31].

The reconstruction stage of SGWT is a reverse procedure of the decomposition stage, which includes inverse update step, inverse prediction step and merging step. In the reconstruction stage, the subband coefficients to be reconstructed are reserved, and then other subband coefficients are set to be zeroes. Finally, the reconstructed results are obtained by the following formula:

$$\begin{cases} X_{l,2^l e} = X_{l+1,2^{l+1}-1} - U(X_{l+1,2^{l+1}}) \\ X_{l,2^l o} = X_{l+1,2^{l+1}} + P(X_{l,2^l e}) \\ X_{l,2^l}(2i) = X_{l,2^l e} \\ X_{l,2^l}(2i+1) = X_{l,2^l o} \\ \cdots \\ X_{l,1e} = X_{l+1,1} - U(X_{l+1,2}) \\ X_{l,1o} = X_{l+1,2} + P(X_{l,1e}) \\ X_{l,1}(2i) = X_{l,1e} \\ X_{l,1}(2i+1) = X_{l,1o} \end{cases} \tag{3}$$

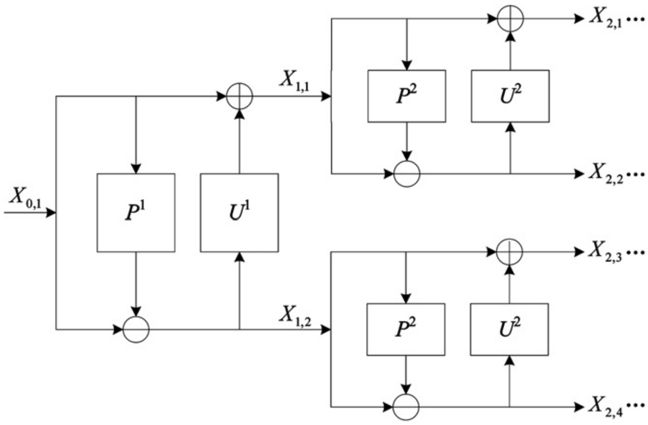

Fig. 2. Decomposition stage of RSGWPT.

Secondly, in the redundant lifting scheme, the split step is discarded. Assuming P^l and U^l, respectively, represent the prediction and update operator of the redundant lifting scheme at level l, the coefficients of P^l and U^l are obtained by padding P_i and U_j of initial operator P and U with zeroes [11].

$$p_i^l = p_0^0,\underbrace{0,\ldots,0}_{2^l-1},p_1^0,\underbrace{0,\ldots,0}_{2^l-1},p_2^0,\ldots,p_{M-2}^0\underbrace{0,\ldots,0}_{2^l-1},p_{M-1}^0 \tag{4}$$

$$u_j^l = u_0^0,\underbrace{0,\ldots,0}_{2^l-1},u_1^0,\underbrace{0,\ldots,0}_{2^l-1},u_2^0,\ldots,u_{N-2}^0\underbrace{0,\ldots,0}_{2^l-1},u_{N-1}^0 \tag{5}$$

With the redundant lifting scheme and SGWPT, the RSGWPT is easily to be constructed. The prediction step and update step of RSGWPT at level l are performed using P^l and U^l, which are expressed as follows:

$$\begin{cases} X_{l+1,2} = X_{l,1} - P^{l+1}(X_{l,1}) \\ X_{l+1,1} = X_{l,1} + U^{l+1}(X_{l+1,2}) \\ \cdots \\ X_{l+1,2^{l+1}} = X_{l,2^l} - P^{l+1}(X_{l,2^l}) \\ X_{l+1,2^{l+1}-1} = X_{l,2^l} + U^{l+1}(X_{l+1,2^{l+1}}) \end{cases} \tag{6}$$

The reconstruction stage of RSGWPT can be obtained from its decomposition stage and expressed by following equations:

$$\begin{cases} X_{l,2^l} = (1/2)(X_{l+1,2^{l+1}-1} - U^{l+1}(X_{l+1,2^{l+1}}) + X_{l+1,2^{l+1}} + P^{l+1}(X_{l+1,2^{l+1}-1} - U^{l+1}(X_{l+1,2^{l+1}}))) \\ \cdots \\ X_{l,1} = (1/2)(X_{l+1,1} - U^{l+1}(X_{l+1,2}) + X_{l+1,2} + P^{l+1}(X_{l+1,1} - U^{l+1}(X_{l+1,2}))) \end{cases} \tag{7}$$

The decomposition and reconstruction stages of RSGWPT are shown in Figs. 2 and 3.

Owing to without the split operation in the decomposition stage of RSGWPT, the approximation and detail coefficients at all levels have the same length as that of the input signal. Consequently, the decomposition results of RSGWPT possess time invariant property and keep the information of the raw signal perfectly. Moreover, it also has inhibiting frequency aliasing property [14], this is very useful for fault feature extraction and fault diagnosis.

After the sampled vibration data are decomposed using RSGWPT, from each of the resultant sub-band wavelet packet coefficients, the nine statistical characteristics including peak value, mean, standard deviation, root mean square, shape factor, skewness, kurtosis, crest factor and pulse index are calculated to describe the distribution property of the signal at each scale.

If the decomposition levels of RSGWPT is l, i.e., the number of sub-band wavelet packet is 2^l, and n statistical characteristics are extracted from each sub-band, then the formed faulty feature vector has $2^l \times n$ feature parameters. According to these statistical characteristics, the feature vector $T=\{f_1, f_2, \ldots, f_{2^l\times n}\}$ can be obtained and the mapping relationship of the feature vector with faulty causes and category can also be constructed.

3. NRS and faulty feature selection

The motivation of rough set based feature selection is to select a minimal attribute subset, which has the same characterizing power as the whole attribute set. Classical rough set model is only suitable for processing the categorical

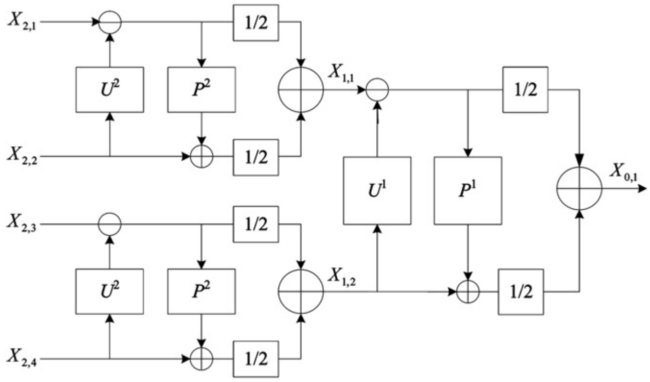

Fig. 3. Reconstruction stage of RSGWPT.

feature. While for the numerical data, which extensively exist in the reality application of mechanical fault diagnosis, it is necessary to convert numerical attributes to categorical attributes. The researchers usually adopt the discretizing algorithm to partition the value domains of real-valued variables into several intervals and then regard them as categorical features. This conversion may not only cause the information loss, but also makes the obtained results depending on the discretization effect in a great degree.

For solving this problem, a neighborhood rough set model was proposed [20,21]. The basic concept and principle are introduced in the following.

In the information system $\langle U,A\rangle$, where U is a non-empty and finite set of samples $\{x_1,x_2,\ \ldots,\ x_n\}$ that is called a universe, A is a set of attributes $\{a_1,\ a_2,\ \ldots,\ a_m\}$ (also called features), which characterize the samples. Specifically, $\langle U,A\rangle$ is also called a decision table if $A=C\cup D$, where C is the set of condition attributes and D is the decision attribute.

Given $x_i\in U$, $B\subseteq C$, the neighborhood of x_i in feature space B is defined as

$$\delta_B(x_i)=\{x_j\,|\,x_j\in U,\Delta_B(x_i,x_j)\le\delta\} \tag{8}$$

where δ is the neighborhood size and Δ is a distance function. Specifically, if B is a categorical feature set then $\delta=0$. The general distance function called the Minkowsky distance is defined as

$$\Delta_B(x_i,x_j)=\left[\sum_{k=1}^{n}(x_{ik}-x_{jk})^P\right]^{1/P} \tag{9}$$

where x_{ik} and x_{jk} are the values of feature k for samples x_i and x_j, respectively. P is a real number that is not less than 1. There are three kinds of distance function according to the selected different P values: (1) Manhattan's distance Δ_1 if $P=1$; (2) Euclidean's distance Δ_2 if $P=2$; (3) Chebychev's distance Δ_3 if $P=\infty$.

$\delta_B(x_i)$ is the neighborhood information granule centered with sample x_i, which is influenced by two important factors. One is the used distance function, which determines the shape of neighborhoods and the other is the threshold δ, which controls the size of the neighborhood.

If an attribute set A contains both numerical attributes B_1 and categorical attributes B_2, the neighborhood of sample x induced by B_1, B_2 and $B_1\cap B_2$ is defined as follows:

$$\delta_{B_1}(x)=\{x_i\,|\,\Delta_{B_1}(x,x_i)\le\delta,x_i\in U\} \tag{10}$$

$$\delta_{B_2}(x)=\{x_i\,|\,\Delta_{B_2}(x,x_i)=0,x_i\in U\} \tag{11}$$

$$\delta_{B_1\cup B_2}(x)=\{x_i\,|\,\Delta_{B_1}(x,x_i)\le\delta\cap\Delta_{B_2}(x_i,x_j)=0\} \tag{12}$$

From the above definition, we can see that the samples in a neighborhood have the same values in terms of categorical features and in terms of numerical features whose distance is less than threshold δ.

A neighborhood relation N on the universe can be written as a relation matrix $(r_{ij})_{n\times n}$, as shown in the Eq. (13). Neighborhood relations draw the objects together for similarity or indistinguishability in terms of distances.

$$r_{ij}=\begin{cases}1, & \Delta(x_i,x_j)\le\delta\\ 0, & \text{otherwise}\end{cases} \tag{13}$$

If there is an attribute in the system generating a neighborhood relation on the universe, then this information system is called a neighborhood information system, denoted by $NIS=\langle U,A,N\rangle$. Specifically, if there are two kinds of attribute sets in the system, i.e. $A=C\cup D$, where C is the condition attributes set and D is the decision attributes set, moreover there exists at least a condition attribute, which includes a neighborhood relation over the universe, a neighborhood information system is also called a neighborhood decision system, which is denoted by $NDT=\langle U,C\cup D,N\rangle$. Suppose $X_1,X_2,\ldots,X_m$ are the object subsets with decision 1–N, $\delta_B(x_i)$ is the neighborhood information of sample x_i generated by a attribute subset $B \subseteq C$. The lower and upper approximations of decision D with respect to attributes B are defined as

$$\underline{N_B}D = \cup_{i=1}^{N} \underline{N_B}X_i, \quad \overline{N_B}D = \cup_{i=1}^{N} \overline{N_B}X_i \tag{14}$$

where

$$\underline{N_B}X = \{x_i \,|\, \delta_B(x_i) \subseteq X, x_i \in U\}, \quad \overline{N_B}X = \{x_i \,|\, \delta_B(x_i) \cap X \neq \emptyset, \quad x_i \in U\} \tag{15}$$

The lower approximation of the decision is also called the positive region of the decision, denoted by $POS_B(D)$. It is the subset of objects whose neighborhood consistently belong to one of the decision classes. On the contrary, the samples in the neighborhood subset of the boundary region ($BN(D)=\underline{N_B}D-\overline{N_B}D$) come from more than one decision class. Thus the neighborhood model divides the samples into the positive region and boundary region of decision. The samples of the positive region can be classified into one of the decision classes without uncertainty, while those of boundary region cannot be determinately classified, accordingly are easy to be misclassified. Therefore, the greater the boundary region is, the weaker the characterizing power of the condition attributes is. For the purpose of characterizing the significance of features, the dependency degree of D–B is defined as

$$\gamma_B(D) = \frac{|POS_B(D)|}{|U|} \tag{16}$$

where $|\bullet|$ is the cardinality of a set. $\gamma_B(D)$ reflects the ability of B to approximate D, which can be used to measure the significance of a feature subset. For $\forall a \in A-B$, the contribution of a single feature a to the approximation of D can be defined as

$$\mathrm{Sig}_B(a,B,D) = \gamma_{B\cup a}(D) - \gamma_B(D) \tag{17}$$

With the above mentioned attribute evaluation method, the forward greedy numerical attribute reduction algorithm based on neighborhood rough set model is employed in this paper [20,21]. Begin with an empty set, the algorithm computes attribute's significance of all the rest of attributes in each round, and the feature which has the maximal attribute's significance is added into the reduction set. The algorithm does not stop until attribute's significance of all the rest of attributes is close to zero. The algorithm is depicted as follows:

- *Step* 1: Given a neighborhood decision system $NDT=\langle U,A,N \rangle$ as input, where U is a nonempty and finite set of samples $\{x_1,x_2, \ldots, x_n\}$; $A=C \cup D$, where C is the set of condition attributes and D is the set of decision attributes;
- *Step* 2: For $\forall a \in A$, compute the neighborhood relation N_a;
- *Step* 3: Let $\emptyset \rightarrow red$, where red is the pool to contain the selected attributes;
- *Step* 4: The significance of a subset of attributes can be defined by the *SIG* function. For each $a_i \in A-red$, compute $\mathrm{Sig}(a_i,red,D) = \gamma_{red\cup a}(D) - \gamma_{red}(D)$;
- *Step* 5: The aim of attribute selection is to search a subset of attributes such that the classification problem has the maximal consistency in the selected feature spaces, and one of the measures for attribute evaluation is to select the attribute a_k satisfying the following equation $\mathrm{Sig}(a_k,red,D) = \max(\mathrm{Sig}(a_i,red,D))$:
- *Step* 6: If $\mathrm{Sig}(a_k,red,D) > 0$, $red \cup a_k \rightarrow red$, and the procedure go to step (4); else return the reduction result *red*.

4. SVM and fault features classification

The SVM is a statistic machine learning technique that has been widely applied in the pattern recognition area. Let $\{(x_i,y_i), \quad i=1,\ldots,N\}$ be a training sample set S and each sample x_i belongs to a class by $y_i \in \{-1,1\}$. The goal of SVM is to find a hyperplane which divides S, such that all the points with the same label are on the same side of the hyperplane while maximizing the distance between the two classes A, B and the hyperplane. An example of the optimal hyperplane of two data sets is presented in Fig. 4.

As shown in Fig. 4, rings and diamonds stand for these two classes of sample points, respectively; H is a separating plane. H_1 and H_2 are the planes that are parallel to H and, respectively, pass through the sample points closest to H in these two classes. The distance between H_1 and H_2 is defined as margin. The optimal separating plane that has the smallest generalization error is the one that not only correctly separates all sample points into these two classes but also leaves the largest margin between H_1 and H_2. SVM can be used in nonlinear classification tasks with application of kernel functions. The basic idea is to transform input vectors into a high dimensional feature space using a nonlinear transformation, and then to do a linear separation in feature space. In order to construct a nonlinear support vector classifier, the inner product (x_i,x) is replaced by a kernel function $K(x_i,x)$, as shown in the following equation:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i,x) + b\right) \tag{18}$$

The SVM has two layers. During the learning process, the first layer selects the basis $K(x_i, x)$, $i=1, 2,\ldots, N$, from the given set of bases defined by the kernel; the second layer constructs a linear function in this space. This is completely equivalent

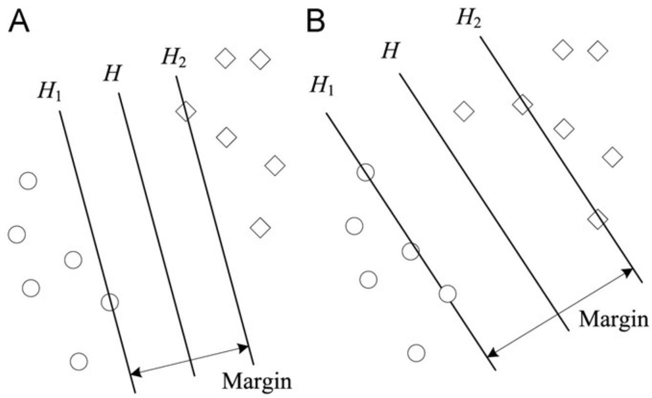

Fig. 4. (A) A separating plane with small margin and (B) a separating plane with larger margin.

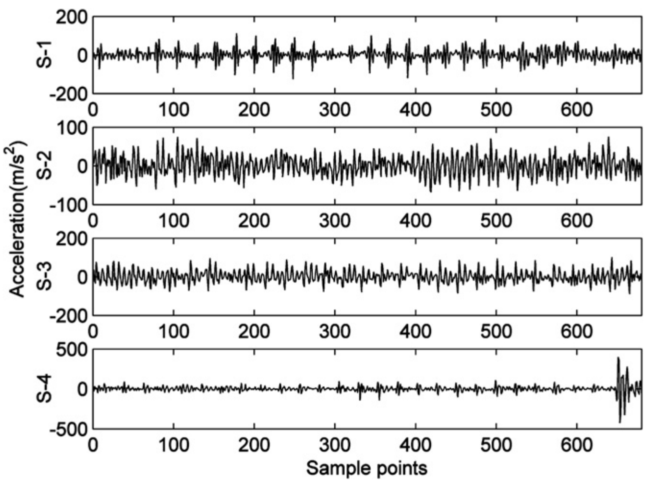

Fig. 5. Gearbox vibration signals in one period of four conditions.

to constructing the optimal hyperplane in the corresponding feature space. The SVM algorithm can construct a variety of learning machines by use of different kernel functions such as linear, polynomial, Gaussian and Laplacian radial basis function.

In this paper, the sequential minimal optimization (SMO) algorithm was adopted for training a support vector classifier. The radial basis function kernel is selected and the complexity parameter C was set to 10. All algorithm are tested by ten fold cross validation, which divided samples into ten folds and then 9 folds are used for training and the remaining 1 fold for testing. The results of the 10 tests are given as a mean of all tests.

5. Experimental setup and data acquisition

To demonstrate the performance of the proposed fault diagnosis method, two kinds of experimental setups are presented in this section for offering the vibration signals from various fault conditions of gearbox and valve trains on a gasoline engine.

5.1. The gearbox experiment

The experimental setup consists of a four-speed motorcycle gearbox, an electrical motor with a constant nominal rotation speed of 1420 rpm, a load mechanism, multi-channel pulse analyzer system, a triaxial accelerometer, tachometer and four shock absorbers under the base of test-bed [32]. The vibration signals measured from the experimental testing of gearbox using the accelerometer, which was mounted on the outer surface of the bearing case of input shaft of the gearbox. The measured vibration signals is dimensionally synchronized using piecewise cubic hermite interpolation method from the revolution point of view, which were not equal following signals acquired by tachometer. After the synchronization of signals per each revolution, the vibration data in one period of these four states are shown in Fig. 5.

Table 1
Five states of valve train on gasoline engine.

State code	Faulty description
S-1	Normal state
S-2	Intake valve clearance of the 4th cylinder and exhaust valve clearance of the 4th cylinder are too large
S-3	Intake valve clearance of the 3rd and 4th cylinder and exhaust valve clearance of the 2nd cylinder are too large
S-4	Intake valve clearance of the 4th and exhaust valve clearance of the 1st cylinder are too large
S-5	Intake valve clearance of the 3rd and 4th and exhaust valve clearance of the 1st and 2nd cylinder are too large

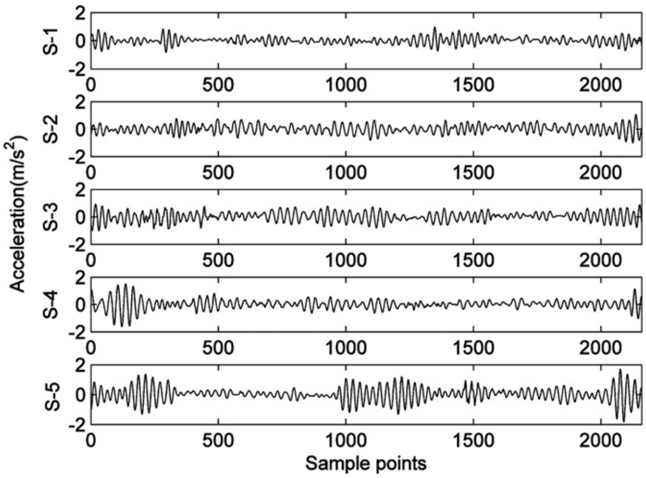

Fig. 6. Valve train vibration signals in one period of five conditions.

In Fig. 5, S-1, S-2, S-3 and S-4 represent faultless condition, slight-worn condition, medium-worn condition and broken-teeth condition, respectively.

5.2. The gasoline engine valve trains experiment

The second example is to identify different conditions of valve train on a gasoline engine. The dead line of the fourth cylinder generated by the pulse sensor is set as the basis point for all the signals. A working cycle of vibration signals are recorded, which include about 720° of crank angle. The vibration acceleration sensor is mounted on the cylinder head and the sampling frequency is 16 kHz. Several typical faults of the valve train are simulated, such as the intake valve clearance being too large and the exhaust valve clearance being too large. Five states including four faulty states and one normal state of valve train are tested. The description of the five states of valve train is shown in Table 1. The vibration data in one cycle of these five states are shown in Fig. 6.

6. Results and discussion

In order to identify different working conditions of the monitoring gearbox and valve train on gasoline engine, the proposed fault diagnosis method mentioned above is performed.

6.1. Case 1: fault diagnosis of gearbox

Firstly, for the fault diagnosis of gearbox, the sample data is decomposed to level 3 using RSGWPT, and the initial operators P and U with $M=4$ and $N=4$, respectively, are chosen. Based on the ISM, the initial prediction operator P is $[-0.0625, 0.5625, 0.5625, -0.0625]$ and the initial update operator U is $[-0.0313, 0.2813, 0.2813, -0.0313]$. Then eight sub-band coefficients are obtained and consequently seventy-two statistical characteristics for each sample can be calculated. According to the faulty features classification method described in Section 5, the classification accuracies using all faulty features and using the selected faulty features through NRS are investigated. The results are listed in Table 2.

The classification accuracies of gearbox experiment are almost the same for the method using all faulty features and that using the selected faulty features through NRS, because this experimental data have lower classification complexity. For the purpose of interpreting the advantage of the proposed faulty diagnosis method, the significance of single faulty features calculated by neighborhood rough set dependency function is shown in Fig. 7, and the size of neighborhood is set to 0.2.

Table 2
Classification accuracies (%) of gearbox experiment obtained using all faulty features and using the selected faulty features through NRS.

Decomposition level	Features extracted	Features selected	Classification accuracies using all features	Classification accuracies using selected features
1	18	4	99.94	99.96
2	36	3	99.98	100
3	72	2	100	100

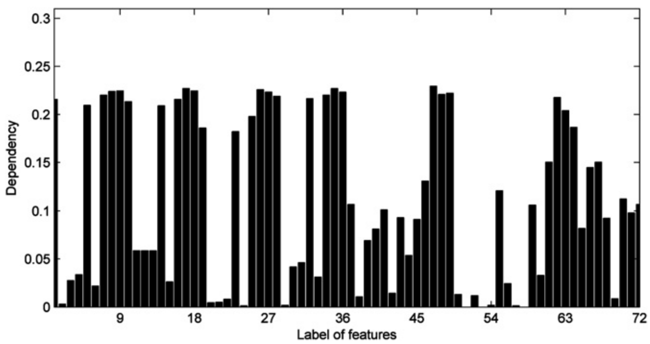

Fig. 7. Significance of single features with neighborhood dependency of gearbox experiment.

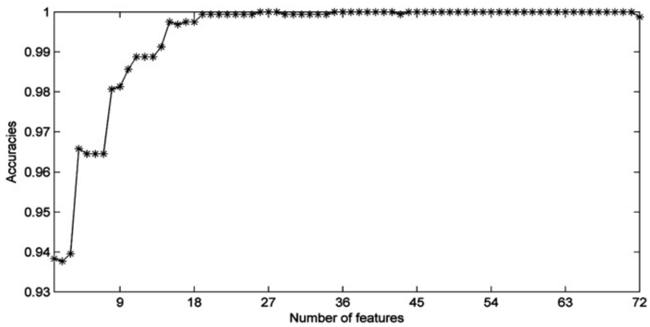

Fig. 8. Accuracies increasing with features added in the significance descending order of gearbox experiment.

The results show that the significances of extracted faulty features are different. Some features' significance is larger, and some is zero. Consequently, the features in the faulty feature vector are not all useful for identifying different working conditions, that is, it contains some useless and redundant information. In order to explain this problem, the features are sorted descending according to their significance, and then successively use one of the best, two of the best ,…, k of the best features to construct the classification model. Fig. 8 shows that the classification accuracies with features added in the descending order of significance using this method.

From Fig. 8, we can find that, for the gearbox experiment, the classification accuracies are improving with the increasing in the number of adding features. But when the number is larger than 20, the classification accuracies stop improving, which means that the latter faulty features is completely useless for faulty pattern classification. This method is called as the sort feature selection method. The existing problem in this method is not capable of identifying the redundant features. The greedy search strategy mentioned in Section 3 can alleviate this disadvantage to some extent. Fig. 9 shows that the variation of classification accuracies with number of selected features using the forward greedy numerical attribute reduction algorithm based on NRS. We can find that using the feature selection based on greedy search strategy, for the gearbox experiment only need approximately 2 features make the classification accuracies achieve more than 99%. It appears to be that a lot of features in the faulty feature vector are able to be deleted, accordingly improve the learning speed and classifier efficiency.

6.2. Case 2: fault diagnosis of gasoline engine valve trains

For the fault diagnosis of the valve train on gasoline engine, because the differences among these five states are very small, that is to say, the classification complexities of different states are higher compared to the sample data of the gearbox experiment, so the experimental results can show the advantage of the proposed diagnosis method more clearly.

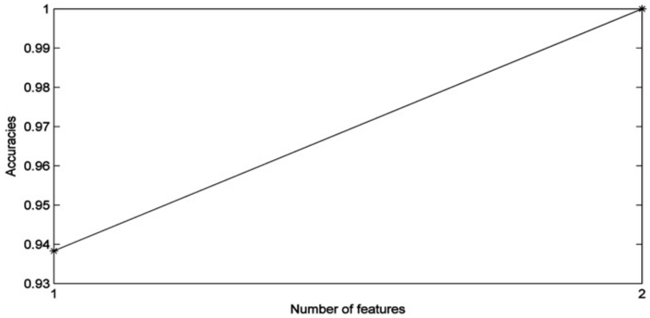

Fig. 9. Variation of accuracies with number of selected features using the forward greedy numerical attribute reduction algorithm based on NRS of gearbox experiment.

Table 3
Classification accuracies (%) of valve train on gasoline engine obtained using all faulty features and using the selected faulty features through NRS.

Decomposition level	Features extracted	Features selected	Classification accuracies using all features	Classification accuracies using selected features
1	18	14	82.37	83.10
2	36	18	89.13	89.16
3	72	13	94.48	94.41
4	144	11	97.34	97.28
5	288	9	97.83	98.01

The sample data is decomposed to level 5 using RSGWPT, and the initial operators P and U are chosen as the same as the above gearbox experiment. The results are listed in Table 3.

From Table 3, we can see that the classification accuracies obtained using the selected faulty features through NRS are close to or little more than those obtained using all faulty features. But considering the time and space cost in the pattern classification, the proposed diagnosis method containing feature selection through NRS is obviously a comprehensive performance optimization method. Additionally, the results can also show that the classification accuracies are increasing with the decomposition level and at the same time the increase rates tends to be reduced. It appears to be that the increase of decomposition level will lead to higher classification performance. On the other hand, the computation cost of RSGWPT and the complexity of classification are increasing with the decomposition level. So when selecting the decomposition level for a specific fault diagnosis problem, there is an unavoidable tradeoff between classification accuracy and the complexity of algorithm.

The size of neighborhood is set to 0.2, and the significance of single faulty features at levels 3 and 5 calculated by neighborhood rough set dependency function are, respectively, shown in Fig. 10(A) and (B). The results again show that the features in the faulty feature vector are not all useful for identifying different working conditions and it contains redundant information.

Figs. 11 and 12 show that at levels 3 and 5 the classification accuracies with features added in the descending order of significance and the variation of classification accuracies with number of selected features using the forward greedy numerical attribute reduction algorithm based on NRS.

From Figs. 11 and 12, we can find that the classification accuracies are improving with the increasing in the number of adding features until the number is 27 (at level 3) and 72 (at level 5), the classification accuracies stop improving. While for the method using the forward greedy numerical attribute reduction algorithm based on NRS, only need approximately 13 (at level 3) and 9 (at level 5) features make the classification accuracies achieve more than 94% and 98%, respectively.

6.3. Discussion

According to [32], for the fault diagnosis of gearbox, firstly the sample data is decomposed to level 4 using WPT, and sixteen subband coefficients are obtained and consequently standard deviation values at each subband can be calculated to constitute faulty feature vector. Thus there are sixteen faulty feature parameters in the feature vector. Lastly, the feature vector is input into artificial neural network to accomplish faulty pattern classification. The results show that using this faulty diagnosis method the classification accuracies can achieve 100%. But in this paper, the experimental results of the

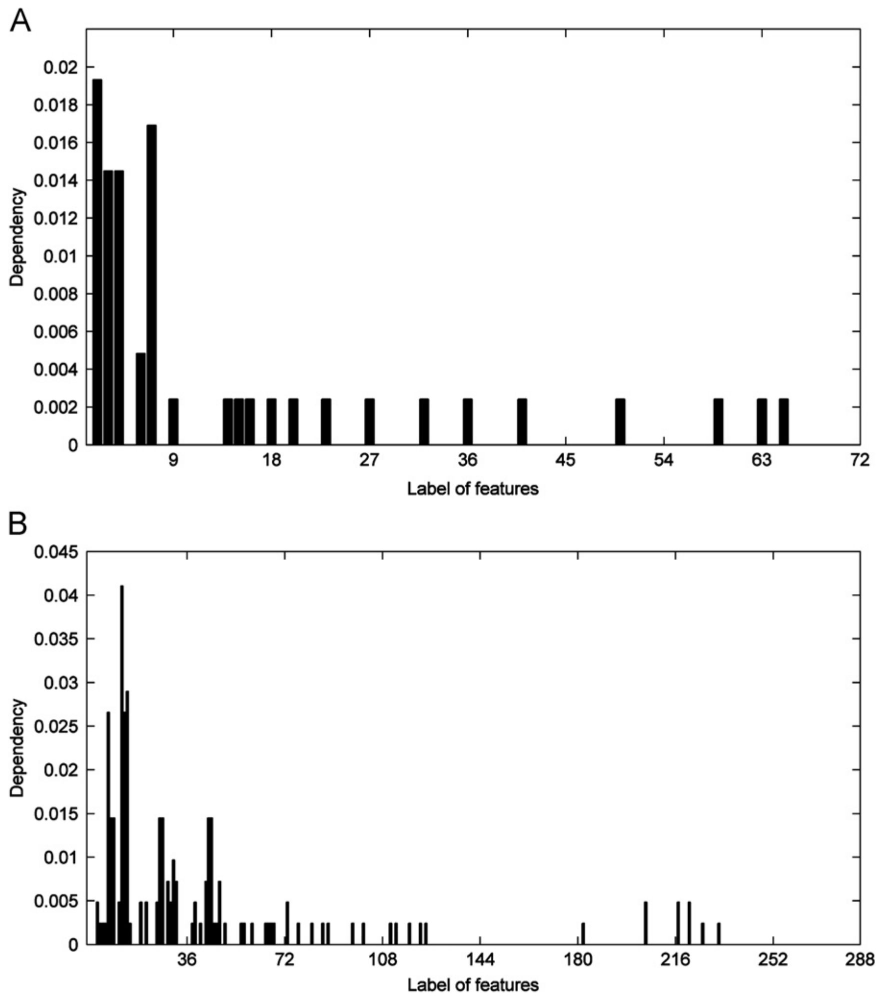

Fig. 10. Significance of single features with neighborhood dependency of engine valve trains experiment. (A) Significance of single features with neighborhood dependency at level 3 and (B) significance of single features with neighborhood dependency at level 5.

proposed method to fault diagnosis of the gearbox (the same vibration signal samples as that in [32]), the classification accuracies can achieve 100% only using 2 faulty feature parameters. For the gasoline engine valve trains experiment, the situation is the same as the first sample. The experimental results show that neighborhood-based feature selection algorithm is able to delete most of redundant and irrelevant features, and the proposed faulty diagnosis method based on RSGWPT, NRS and SVM can extract the faulty features, which have better classification ability and at the same time reduce a lot of redundant features in case of assuring the classification accuracy, accordingly improve the classifier efficiency and achieve a better classification performance.

There are some fault feature selection methods such as distance evaluation technique [33,34], genetic algorithm [23,35], rough set [5,15–18], etc. Due to the simpleness and reliability of the distance evaluation technique, it is generally adopted in fault diagnosis recently [3,4]. The distance evaluation technique can evaluate the significance of single faulty feature by feature evaluation indexes, but it cannot obtain the features combination to directly achieve the best classification performance. For genetic algorithm, it is essential to determine the fitness function, which is directly related to the velocity of convergence and finding of the optimum features subset. Unfortunately, the selection of fitness function and correlated parameters depends on the prior knowledge and experiences. The classical rough set model can just be used to process categorical features that are discrete values, so for the rough set based feature selection in mechanical fault diagnosis system, it is necessary to introduce a discretizing algorithm to partition the value domains of real-valued variables into several intervals. Obviously, discretization of numerical attributes may lead to the loss of classification information [21,22]. Compared with the above mentioned methods, the proposed faulty feature selection method based on neighborhood rough set model in this paper has the advantage of perfectly preserving the fault information of original data and obtaining the optimum features subset, at the same time it does not need any prior knowledge.

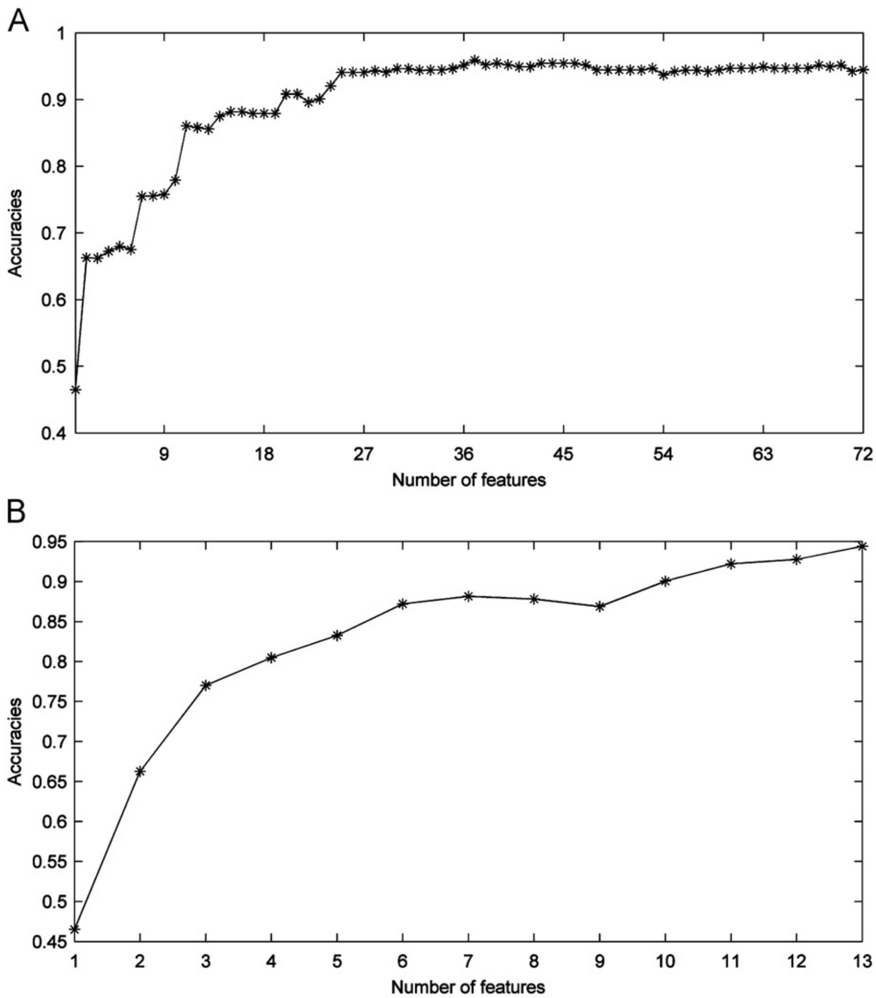

Fig. 11. Variation of accuracies of engine valve trains experiment at level 3. (A) Accuracies increasing with features added in the significance descending order and (B) variation of accuracies with number of selected features using the forward greedy numerical attribute reduction algorithm based on NRS.

The purpose of this paper is to afford a novel relatively generalized model for mechanical equipment faulty intelligent diagnosis. Above, the proposed method was applied to diagnose the faults of the gearbox (a typical rotating machinery) and the gasoline engine (a typical reciprocating machinery) valve trains. In the gearbox experiment, from the signals illustrated in Fig. 5, we can obviously find that the experimental data have lower classification complexity. In the gasoline engine valve trains experiment, the differences among the signals in Fig. 6 are insignificant, therefore the classification complexities are higher than the first case. The experimental results show that the proposed faulty diagnosis method are effective for both simple and complex mechanical diagnosis problem. The difference is that for the relatively complex faulty diagnosis problem, such as the gasoline engine valve trains experiment, more faulty feature parameters are retained. Thus we can conclude that the number of faulty features are increasing with the complexity of fault classification problem in case of assuring the classification accuracy. It is worth mentioning that this difference is determined by the property of faulty diagnosis model without requiring human intervention and hence this method has the self-adaptive characteristics to some extent.

7. Conclusion

In this paper, we have proposed a novel method for mechanical faulty diagnosis based on RSGWPT, NRS and SVM. This method utilizes the RSGWPT to extract faulty feature parameters from the statistical characteristics of wavelet package coefficients to constitute feature vectors, and then makes the attribute reduction through NRS method to obtain the key features. These key features have better distinction ability which is pivotal to classification, so it is more effective to input the key features other than all features into SVM to accomplish faulty pattern classification. The approach can select an optimal subset of features

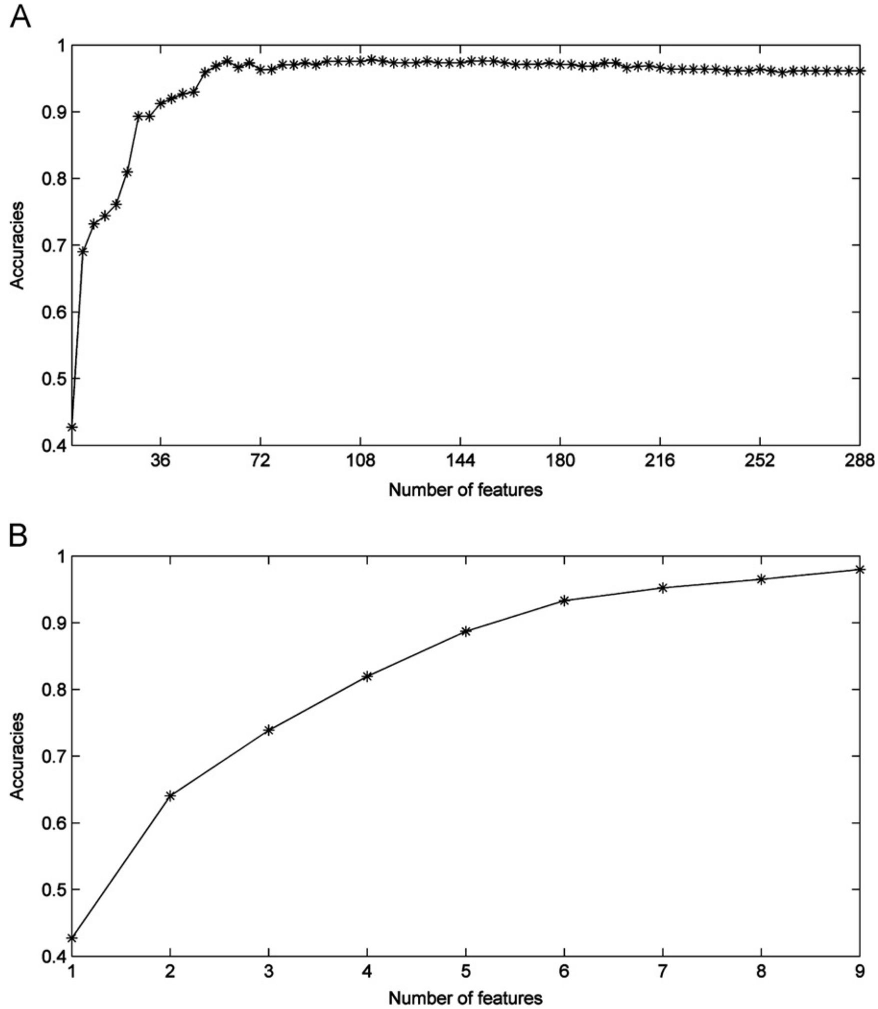

Fig. 12. Variation of accuracies of engine valve trains experiment at level 5. (A) Accuracies increasing with features added in the significance descending order and (B) variation of accuracies with number of selected features using the forward greedy numerical attribute reduction algorithm based on NRS.

quickly and effectively from a large database with a lot of features. The validity of the proposed fault diagnosis method is verified by the application of practical experiments of the gearbox and gasoline engine valve trains.

Acknowledgments

This work is supported by School Foundation of Shanghai Second Polytechnic University (Project no. QD210005) and Leading Academic Discipline Project of Shanghai Municipal Education Commission (Project no. J51802).

References

[1] Z.K. Peng, F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mech. Syst. Signal Process. 18 (2) (2004) 199–221.
[2] C.Z. Chen, C.T. Mo, A method for intelligent fault diagnosis of rotating machinery, Digital Signal Process. 14 (3) (2004) 203–217.
[3] Y.G. Lei, Z.J. He, Y.Y. Zi, A new approach to intelligent fault diagnosis of rotating machinery, Expert Syst. Appl. 35 (4) (2008) 1593–1600.
[4] Q. Hu, Z.J. He, Z.S. Zhang, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble, Mech. Syst. Signal Process. 21 (2) (2007) 688–705.
[5] X.Q. Xiang, J.Z. Zhou, C.S. Li, et al., Fault diagnosis based on Walsh transform and rough sets, Mech. Syst. Signal Process. 23 (4) (2009) 1313–1326.
[6] W. Sweldens, The lifting scheme: a custom-design construction of biorthogonal wavelets, Appl. Comput. Harmonic Anal. 3 (2) (1996) 186–200.
[7] W. Sweldens, The lifting scheme: a construction of second generation wavelets, SIAM J. Math. Anal. 29 (2) (1998) 511–546.
[8] H.K. Jiang, Z.S. Wang, Second generation wavelet packet construction and aircraft engine weak damage identification, J. Beijing Univ. Aeronaut. Astronaut. 33 (7) (2007) 777–780 (in Chinese).
[9] J.F. Leng, D.H. Chen, S.X. Jing, Mine fan intelligent faults diagnosis based on the lifting wavelet packet and RBF neural network, in: Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery, Haikou, China, 2007, vol. 4, pp. 716–720.

[10] Y.N. Pan, J. Chen, L. Guo, Robust bearing performance degradation assessment method based on improved wavelet packet-support vector data description, Mech. Syst. Signal Process. 23 (3) (2009) 669–681.

[11] C.S. Lee, C.K. Lee, K.Y. Yoo, New lifting based structure for undecimated wavelet transform, Electron. Lett. 36 (22) (2000) 1894–1895.

[12] R.L. Claypoole, Adaptive Wavelet Transform via Lifting, Ph.D. Thesis, Department of Electrical and Computer Engineering, Rice University, Houston, Texas, 1999.

[13] Z. Li, Z.J. He, Y.Y. Zi, X.F. Chen, Bearing condition monitoring based on shock pulse method and improved redundant lifting scheme, Math. Comput. Simulation 7 (3) (2008) 318–338.

[14] W. Bao, R. Zhou, J.G. Yang, D.R. Yu, N. Li, Anti-aliasing lifting scheme for mechanical vibration fault feature extraction, Mech. Syst. Signal Process. 23 (5) (2009) 1458–1473.

[15] F.E.H. Tay, L.X. Shen, Fault diagnosis based on rough set theory, Eng. Appl. Artif. Intell. 16 (1) (2003) 39–43.

[16] R. Zhou, J.G. Yang, The research of engine fault diagnosis based on rough sets and support vector machine, Trans. CSICE 24 (4) (2006) 379–383 (in Chinese).

[17] J.R. Li, L.P. Khoo, S.B. Tor, RMINE: a rough set based data mining prototype for the reasoning of incomplete data in condition-based fault diagnosis, J. Intell. Manuf. 17 (1) (2006) 163–176.

[18] Z.Q. Geng, Q.X. Zhu, Rough set-based heuristic hybrid recognizer and its application in fault diagnosis, Expert Syst. Appl. 36 (2) (2009) 2711–2718.

[19] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches, IEEE Trans. of Knowl. Data Eng. 16 (12) (2004) 1457–1471.

[20] Q.H. Hu, D.R. Yu, J.F. Liu, et al., Neighborhood rough set based heterogeneous feature subset selection, Inform. Sci. 178 (18) (2008) 3577–3594.

[21] Q.H. Hu, D.R. Yu, Z.X. Xie, Neighborhood classifiers, Expert Syst. Appl. 34 (2) (2008) 866–876.

[22] Q.H. Hu, J.F. Liu, D.R. Yu, Mixed feature selection based on granulation and approximation, Knowledge-Based Syst. 21 (4) (2008) 294–304.

[23] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mech. Syst. Signal Process. 18 (3) (2004) 625–644.

[24] B.S. Yanga, W.W. Hwang, D.J. Kim, A.C. Tan, Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines, Mech. Syst. Signal Process. 19 (2) (2005) 371–390.

[25] V. Sugumaran, V. Muralidharan, K.I. Ramachandran, Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing, Mech. Syst. Signal Process. 21 (2) (2007) 930–942.

[26] S.F. Yuan, F.L. Chu, Fault diagnosis based on support vector machines with parameter optimisation by artificial immunisation algorithm, Mech. Syst. Signal Process. 21 (3) (2007) 1318–1330.

[27] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.

[28] M. Ge, R. Du, G. Zhang, et al., Fault diagnosis using support vector machine with an application in sheet metal stamping operations, Mech. Syst. Signal Process. 18 (1) (2004) 143–159.

[29] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal Process. 21 (6) (2007) 2560–2574.

[30] R.L. Claypoole, G.M. Davis, W. Sweldens, R. Baraniuk, Nonlinear wavelet transforms for image coding via lifting, IEEE Trans. Image Process 12 (12) (2003) 1449–1459.

[31] J. Stepien, T. Zielinski, R. Rumian, Image denoising using scale-adaptive lifting schemes, in: Proceedings of the International Conference on Image, Vancouver, BC, Canada, 2000, vol. 3, pp. 288–290.

[32] J. Rafiee, F. Arvani, A. Harifi, M.H. Sadeghi, Intelligent condition monitoring of a gearbox using artificial neural network, Mech. Syst. Signal Process. 21 (4) (2007) 1746–1754.

[33] B.S. Yang, T. Han, J.L. An, ART-KOHONEN neural network for fault diagnosis of rotating machinery, Mech. Syst. Signal Process. 18 (2004) 645–657.

[34] B.S. Yang, T. Han, W.W. Hwang, Fault diagnosis of rotating machinery based on multi-class support vector machines, J. Mech. Sci. Technol. 19 (3) (2005) 846–859.

[35] L. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms, Mech. Syst. Signal Process. 16 (2002) 373–390.