

Research Article

Feature Extraction and Selection Scheme for Intelligent Engine Fault Diagnosis Based on 2DNMF, Mutual Information, and NSGA-II

Peng-yuan Liu, Bing Li, Cui-e Han, and Feng Wang

Forth Department, Mechanical Engineering College, No. 97, Heping West Road, Shijiazhuang, Hebei 050003, China

Correspondence should be addressed to Bing Li; rommandy@163.com

Received 24 December 2015; Revised 20 March 2016; Accepted 27 March 2016

Academic Editor: Arturo Garcia-Perez

Copyright © 2016 Peng-yuan Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel feature extraction and selection scheme is presented for intelligent engine fault diagnosis by utilizing two-dimensional nonnegative matrix factorization (2DNMF), mutual information, and nondominated sorting genetic algorithms II (NSGA-II). Experiments are conducted on an engine test rig, in which eight different engine operating conditions including one normal condition and seven fault conditions are simulated, to evaluate the presented feature extraction and selection scheme. In the phase of feature extraction, the *S* transform technique is firstly utilized to convert the engine vibration signals to time-frequency domain, which can provide richer information on engine operating conditions. Then a novel feature extraction technique, named two-dimensional nonnegative matrix factorization, is employed for characterizing the time-frequency representations. In the feature selection phase, a hybrid filter and wrapper scheme based on mutual information and NSGA-II is utilized to acquire a compact feature subset for engine fault diagnosis. Experimental results by adopted three different classifiers have demonstrated that the proposed feature extraction and selection scheme can achieve a very satisfying classification performance with fewer features for engine fault diagnosis.

1. Introduction

Engine is one of the core mechanical components in a wide range of industrial applications. Detecting and identifying the faults of engine at an early stage while the machine is still on operation can help to avoid serious accidents and reduce economic losses. For this reason, various intelligent techniques such as artificial neural networks (ANN), support vector machine (SVM), relevance vector machine (RVM), and extreme learning machines (ELM) have been successfully applied to automated detection and diagnosis of engine conditions over the past few years [1–6].

For any intelligent fault diagnosis systems, feature extraction and feature selection schemes can be regarded as the two most important steps. Feature extraction is a mapping process from the measured signal space to the feature space. Representative features associated with the conditions of machinery components should be extracted by using appropriate signal processing and calculating approaches. Various

techniques, such as wavelet analysis, empirical mode decomposition, multivariate statistics, morphological pattern spectrum, fractal theory, and nonnegative matrix factorization, have been applied in engine fault diagnosis in the past few years [7–11]. In this work, a new feature extraction approach based on *S* transform [12] and two-dimensional nonnegative matrix factorization (2DNMF), which has been used for bearing fault diagnosis in our earlier work [13], is employed for engine fault diagnosis. To the best of our knowledge, this feature extraction has never been utilized in engine fault diagnosis. The main advantage of 2DNMF to NMF is that it does not need to transform the 2D matrix to 1D vector, which will lead to a high-dimensional vector space and loss of some structure information hiding in original 2D matrices. Thus, the 2DNMF is capable of characterizing the time-frequency representations more effectively and efficiently for engine fault diagnosis.

Filter and wrapper methods can be regarded as the two main categories of feature selection approaches in the

literatures. Several researches have been done by utilizing the filter method or wrapper method for mechanical fault diagnosis utilizing Fisher's criterion [14], distance evaluation technique [15, 16], decision tree [17, 18], and evolution algorithm (EA) combined with ANN and SVM [19–24]. They have largely improved the efficiency and accuracy of mechanical fault diagnosis in practice. Despite the successive applications of filter or wrapper methods in fault diagnosis, it should be noted that the wrapper and filter methods can complement each other. The filter methods can search through the feature space efficiently but usually fail to obtain a good accuracy, while the wrappers can provide good accuracy but require much computation time. Therefore, it is very desirable to combine the filter and wrapper methods to achieve high efficiency and accuracy simultaneously [25–28]. In our earlier works [29, 30], the hybrid feature selection scheme combining filter and wrapper methods based on ReliefF, genetic algorithm (GA), mutual information, and nondominated sort genetic algorithm (NSGA-II) has been investigated for gear fault diagnosis. However, to the best of our knowledge, these schemes have not yet been used for engine fault diagnosis. Therefore, in this work, the hybrid feature selection scheme based on the mutual information and the improved nondominated sort genetic algorithm (NSGA-II) is employed and verified for engine fault diagnosis. Otherwise, a more detailed demonstration of the above-mentioned feature extraction and selection scheme, which has not been demonstrated in earlier works, is given in this work.

Experiments are conducted on an engine test rig, in which eight different engine operating states including normal state and seven fault states are simulated in the experiments. The performance of the proposed feature extraction and selection scheme is testified on the engine dataset. Experimental results have demonstrated the superiority of the proposed intelligent scheme on computation cost and classification accuracies in engine fault diagnosis. Figure 1 illustrates the flowchart of the intelligent engine fault diagnosis scheme presented in this work.

The remainder of this work is organized as follows. Section 2 describes the feature extraction method based on the S transform and 2DNMF. In Section 3, the hybrid filter and wrapper feature selection scheme based on mutual information and NSGA-II is detailed. Section 4 presents the application results of the proposed feature extraction and selection scheme for engine fault diagnosis. The conclusions of this paper are summarized in Section 5.

2. Feature Extraction Based on S Transform and Two-Dimensional Nonnegative Matrix Factorization

2.1. S Transform. The S transform, put forward by Stockwell et al. in 1996, can be regarded as an extension to the ideas of the Gabor transform and the wavelet transform. The S transform of signal $x(t)$ is defined as

$$S(\tau, f) = \int_{-\infty}^{+\infty} x(t) w(t - \tau) e^{-j2\pi ft} dt, \quad (1)$$

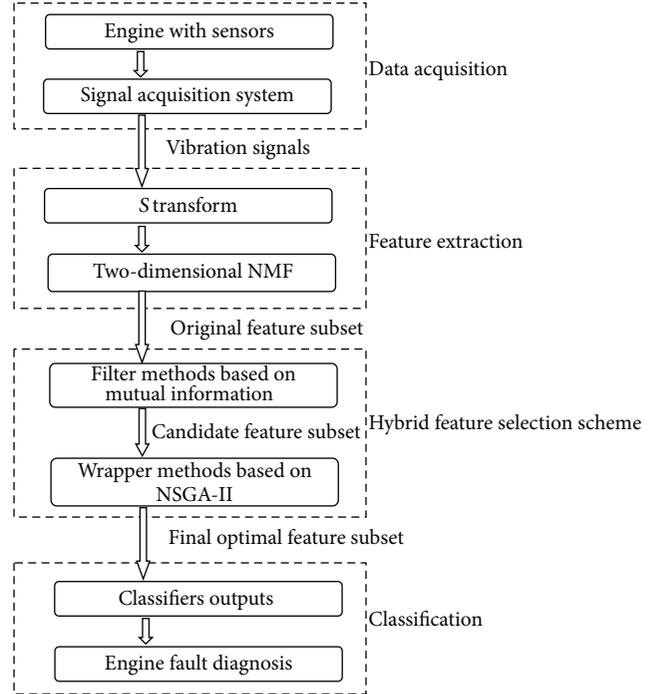


FIGURE 1: Flowchart of the presented feature extraction and selection scheme for intelligent engine fault diagnosis.

where

$$w(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2}, \quad (2)$$

$$\sigma = \frac{1}{|f|}.$$

Then the S transform can be given by combining (1) and (2):

$$S(\tau, f) = \int_{-\infty}^{+\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-(t-\tau)^2 f^2 / 2} e^{-j2\pi ft} dt. \quad (3)$$

Since S transform is a representation of the local spectra, Fourier or time average spectrum can be directly obtained by averaging the local spectra as

$$\int_{-\infty}^{+\infty} S(\tau, f) d\tau = X(f), \quad (4)$$

where $X(f)$ is the Fourier transform of $x(t)$.

The inverse S transform is given by

$$x(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S(\tau, f) e^{j2\pi ft} d\tau df. \quad (5)$$

The main advantage of the S transform over the short-time Fourier transform (STFT) is that the standard deviation σ is actually a function of frequency f . Consequently, the window function is also a function of time and frequency. As the width of the window is controlled by the frequency, it can obviously be seen that the window is wider in the time domain at lower frequencies and narrower at higher

frequencies. In other words, the window provides good localization in the frequency domain for low frequencies while providing good localization in time domain for higher frequencies. It is a very desirable characteristic for accurate representation of nonstationary vibration signals in time-frequency domain.

2.2. Nonnegative Matrix Factorization (NMF)

2.2.1. Nonnegative Matrix Factorization. The NMF algorithm is a technique that compresses a matrix into a smaller number of basis functions and their encodings [31]. The factorization can be expressed as follows:

$$\mathbf{V}_{n \times m} \approx \mathbf{W}_{n \times r} \mathbf{H}_{r \times m}, \quad (6)$$

where \mathbf{V} denotes an $n \times m$ matrix and m is the number of examples in the dataset, each column of which contains an n -dimensional observed data vector with nonnegative values. This matrix then approximately factorized into $n \times r$ matrix \mathbf{W} and $r \times m$ matrix \mathbf{H} . The rank r of the factorization is usually chosen such that $(n + m)r < nm$, and hence the compression or dimensionality reduction is achieved. The key characteristic of NMF is the nonnegativity constraints imposed on the two factors, and the nonnegativity constraints are compatible with the intuitive notion of combining parts to form a whole.

2.2.2. Two-Dimensional Nonnegative Matrix Factorization (2DNMF). The key difference between 2DNMF and NMF is that the former adopt a novel representation for original time-frequency representations. In traditional NMF, the 2D time-frequency matrices must be previously transformed into 1D vector. The resulting vectors usually lead to a high-dimensional vector space, where it is difficult to find good bases to approximately reconstruct original time-frequency distributions. That is also called the ‘‘curse of dimensionality’’ problem, which is more apparent in small-sample-size cases. Another disadvantage of NMF is that such a matrix-to-vector transform may cause the loss of some structure information hiding in original 2D time-frequency representations [13].

Assume $p \times q$ matrix \mathbf{A}_k ($k = 1, 2, \dots, m$) denotes the time-frequency representation of k th training sample signal acquired from engine. In traditional NMF, a 2D time-frequency representation is first transformed into a 1D vector, and then the training databases are represented with $n \times m$ matrix \mathbf{V} , each column of which contains $n = pq$ nonnegative values of one of the m time-frequency representations. In 2DNMF, however, it never transforms the 2D time-frequency representations into their corresponding 1D vector. Instead, it will use a more straightforward way which views a time-frequency representation as a 2D matrix.

The procedure of 2DNMF method consists of two successive stages. At first it aligns m training TFR matrices into a $p \times qm$ matrix $\mathbf{X} = [\mathbf{A}_1 \ \mathbf{A}_2 \ \dots \ \mathbf{A}_m]$, where each \mathbf{A}_k denotes one of the m TFR matrices. Similar to NMF, 2DNMF first finds $p \times d_1$ nonnegative matrix \mathbf{L} and $d_1 \times qm$ nonnegative matrix \mathbf{H} such that

$$\mathbf{X} \approx \mathbf{LH}. \quad (7)$$

Here \mathbf{L} and \mathbf{H} are the basis functions and encoding coefficients, respectively. Since each column of \mathbf{X} corresponds to a column of original TFR matrix, \mathbf{L} is also named as column bases.

The second stage of 2DNMF involves computing the row bases. In this stage, it first constructs a new $q \times pm$ matrix $\mathbf{X}' = [\mathbf{A}'_1 \ \mathbf{A}'_2 \ \dots \ \mathbf{A}'_m]$, where each \mathbf{A}'_k denotes the transpose of \mathbf{A}_k . Similarly, 2DNMF seeks a $q \times d_2$ nonnegative matrix \mathbf{R} and a $d_2 \times pm$ nonnegative matrix \mathbf{C} such that

$$\mathbf{X}' \approx \mathbf{RC}. \quad (8)$$

Here \mathbf{R} and \mathbf{C} are the basis functions and encoding coefficients, respectively.

By now the $p \times d_1$ dimensional column bases \mathbf{L} and the $q \times d_2$ dimensional row bases \mathbf{R} are obtained. A new representation of TFR \mathbf{A}_k projected to the column bases \mathbf{L} and row bases \mathbf{R} can be denoted as

$$\mathbf{D}_k = \mathbf{L}^T \mathbf{A}_k \mathbf{R}, \quad k = 1, 2, \dots, m, \quad (9)$$

where \mathbf{D}_k is a $d_1 \times d_2$ matrix, which can be regarded as a reduced dimension representation of TFR \mathbf{A}_k and can be used as features for fault diagnosis of engine states.

3. Hybrid Filter and Wrapper Feature Selection Scheme Based on Mutual Information and NSGA-II

3.1. Hybrid Filter and Wrapper Feature Selection Scheme. Filters and wrappers are the two main categories of feature selection algorithms in the literatures. Filter methods evaluate the goodness of the feature subset by using the intrinsic characteristic of the data. They are relatively computationally cheap since they do not involve the induction algorithm. However, they also take the risk of selecting subsets of features that may not match the chosen induction algorithm. Wrapper methods, on the contrary, directly use the classifiers to evaluate the feature subsets. They generally outperform filter methods in terms of prediction accuracy, but they are generally computationally more intensive [32–35]. In summary, wrapper and filter methods can complement each other, in that filter methods can search through the feature space efficiently while the wrappers can provide good accuracy. It is desirable to combine the filter and wrapper methods to achieve high efficiency and accuracy simultaneously.

In this work, the hybrid feature selection combining filter and wrapper feature selection technique based on the mutual information and the improved nondominated sort genetic algorithm (NSGA-II) is employed [30]. In this hybrid filter and wrapper feature selection scheme, there are two steps involved. In the first stage, a candidate feature subset is chosen according to the max-relevance and min-redundancy (mRMR) criterion based on mutual information from the original feature set. Then at the second stage, classifier combined with NSGA-II is adopted to find a more compact feature subset from the candidate feature subset. In this stage, feature selection problem is defined as a multiobjective problem dealing with two competing objectives, mean lesser features and lower classification error rate.

3.2. Filter Method Based on Mutual Information

3.2.1. Mutual Information. Mutual information is one of the most widely used measures to define relevancy of variables [36]. In this section, the feature selection method based on mutual information is focused on. Given two random variables x and y , their mutual information can be defined in terms of their probabilistic density functions $p(x)$, $p(y)$, and $p(x, y)$:

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy. \quad (10)$$

The estimation of the mutual information of two variables was detailed in [36].

In supervised classification, one can view the classes as a variable (named C) with L possible values (where L is the number of classes of the system) and the feature component as another variable (named X) with K possible values (where K is the number of parameters of the system). So one will be able to compute the mutual information $I(x_k, c)$ between the classes c and the feature x_k ($k = 1, 2, \dots, K$):

$$I(x_k, c) = \iint p(x_k, c) \log \frac{p(x_k, c)}{p(x_k)p(c)} dx_k dc. \quad (11)$$

Then the informative variables with larger $I(x_k, c)$ can be identified. A more compact feature subset can be obtained via selecting the d best features based on (12) from the original feature set.

Equation (12) provides us with a measure to evaluate the effectiveness of the ‘‘global’’ feature that is simultaneously suitable to differentiate all classes of signals. For a small number of classes, this approach may be sufficient. The more signal classes, the more ambiguous $I(x_k, c)$.

3.2.2. Max-Relevance and Min-Redundancy. Max-relevance means that the selected features x_i are required, individually, to have the largest mutual information $I(x_i, c)$. It means that the m best individual features should be selected according to this criterion. It can be represented as

$$\max D(S, c), \quad (12)$$

$$D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i, c),$$

where $|S|$ denotes the number of features contained by S .

However, it has been proved that the simple combination of the best individual features does not necessarily lead to a good performance. In other words, ‘‘the m best features are not the best m features’’ [36, 37]. The most important problem of the max-relevance is that it neglects the redundancy between features and may cause the degradation of the classification performance.

So the min-redundancy criterion should be added to the selection of the optimal subsets. It can be represented as

$$\min R(S), \quad (13)$$

$$R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j).$$

The criterion combining the above two constraints is called the ‘‘maximal relevance minimal redundancy’’ (mRMR) [36]. The operator $\Phi(D, R)$ is defined to optimize D and R simultaneously:

$$\max \Phi(D, R), \quad (14)$$

$$\Phi = D - R.$$

3.2.3. Candidate Feature Subset Obtained Based on Max-Relevance and Min-Redundancy. In practice, greedy search methods can be used to find the near-optimal features by Φ . Let F be the original feature sets and let S be the selected subsets. Suppose that S already has S_{m-1} , which means it has selected $m - 1$ features. The next work is to select m th feature from the set $\{F - S_{m-1}\}$. This is done according to the following criterion:

$$\max_{x_j \in F - S_{m-1}} \left[I(x_j, c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j, x_i) \right]. \quad (15)$$

The main steps can be represented as follows.

Step 1. Let F be the original feature set and let S be the selected subset. S is initialized to be an empty subset, $S \rightarrow \{\}$.

Step 2. Calculate the relevance of individual feature x_i with the target class c , denoted by $I(x_i, c)$.

Step 3. Find the feature x_k which has the maximum relevance:

$$I(x_k, c) = \max_{x_i \in F} I(x_i, c). \quad (16)$$

Let $F_1 \rightarrow \{F - x_k\}$, $S_1 \rightarrow \{S + x_k\}$.

Step 4. This step consists of the following:

for $m = 2 : N$

Let $x_j \in F_{m-1}$, $x_i \in S_{m-1}$, find the x_k according to the following criterion:

$$\max_{x_j \in F_{m-1}} \left[I(x_j, c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j, x_i) \right] \quad (17)$$

Let $F_m = \{F_{m-1} - x_k\}$, $S_m = \{S_{m-1} + x_k\}$

end

In this way, N sequential feature subsets can be obtained and satisfy $S_1 \subset S_2 \subset \dots \subset S_N$. In practice, the first d features can be selected as the feature subset S_d , which is the candidate feature subset for wrapper methods.

3.3. Wrapper Method Based on NSGA-II

3.3.1. A Brief Review on NSGA-II. The presence of multiple objectives in practical problems has given rise to the rapid

development of multiobjective evolutionary algorithms over the past few years. Nondominated sorting genetic algorithm (NSGA), which was suggested by Goldberg and implemented by Srinivas and Deb [38], has been proved to be an effective approach for multiobjective optimization problems. However, NSGA is still suffering three main drawbacks: the high computational complexity of nondominated sorting, lack of elitism, and requirement for specifying the sharing parameter.

As an improved version of the NSGA, NSGA-II was introduced by Deb et al. in 2002 [39]. The NSGA-II overcame the original NSGA defects by introducing the fast nondominated sorting algorithm to alleviate computational complexity, the elitist-preserving mechanism to speed up the evolution, and the crowded comparison operator to avoid specifying the sharing parameter. It has been verified that the NSGA-II is able to maintain a better spread of solutions and converges better in the obtained nondominated front compared to other similar elitist multiple objectives evolution algorithms (MOEAs). More details about the description and implementation can be found in [39].

3.3.2. Wrapper Feature Selection Using NSGA-II. In most cases of conventional wrapper methods for fault diagnosis, the feature selection problem was formulized as a single objective problem [19, 21, 22, 24, 40]. However, the feature selection is inherently a multiobjective problem, which deals with two competing objectives: the feature dimension and the classification accuracy. An optimal feature set has to be of a minimal number of features and has to produce the minimum classification error rate.

In this work, the feature selection problem for engine fault diagnosis is formulated to be a multiobjective problem. The NSGA-II mentioned above is utilized to optimize the two objectives: the minimal number of features and minimum classification error rate. A step-by-step procedure for solving the feature selection problem by utilizing the NSGA-II is illustrated in Figure 2.

3.3.3. Implementation Issues for Wrapper Feature Selection Using NSGA-II. For wrapper feature selection approach, there are several factors for controlling the process of NSGA-II while searching the suboptimal feature subsets for classifiers. To apply NSGA-II to feature selection, the following issues are focused on.

(1) Fitness Functions. Two competing objectives are defined as the fitness functions: the first was minimization of the number of used features and the second was minimization of the classification error rate. Three popular classifiers, that is, K nearest neighbor classifier (KNNC) [41], Naïve Bayes classifier (NBC) [42], and least-square support vector machine (LS-SVM) [43], are employed as induction algorithms to implement and evaluate the proposed feature selection approach. The KNNC and NMC are implemented by utilizing the MATLAB Toolbox for Pattern Recognition (PRTools 4.1) [44]. The LS-SVM was implemented by the LS-SVMlab1.5, which can be downloaded from [45].

TABLE 1: Engine operating states in experiment.

State code	Engine operating condition
1	Normal state
2	First cylinder misfire
3	Second cylinder misfire
4	First and second cylinder misfires
5	Too large clearance of intake valve
6	Too large clearance of exhaust valve
7	Air leakage of the intake valve
8	Air leakage of the exhaust valve

(2) Encoding Scheme. The binary coding system is used to represent the chromosome in this investigation. For chromosome representing the feature subsets, the bit with value “1” represents the feature being selected, and “0” indicates that feature is not selected, as shown in Figure 3.

(3) Genetic Operators. Genetic operator consists of two basic operators, that is, crossover and mutation. The used crossover technique is the uniform crossover consisting of replacing genetic material of the two selected parents uniformly in several points. The mutation operator used in this work is implemented as conventional mutation operator operating on each bit separately and changing randomly its value.

4. Results and Discussion

4.1. Engine Dataset Description. To evaluate the performance of the presented feature extraction and selection scheme, experiments are carried out on F3L912 engine with three cylinders to measure the vibration signals. Eight engine running conditions, including one healthy engine state and seven faulted states as summarized in Table 1, are tested in the experiments. All the valve defects are set on the first cylinder. The running speed is set to be 1200 rpm.

Vibration data is collected using accelerometers, which are attached to the cylinder head near first cylinder with magnetic bases. The sample frequency is 20 K. A working cycle of vibration signals is recorded as a sample, which include 4096 sample points. Forty samples are recorded for every condition of engine, where there are totally 320 samples. Figure 4 shows the waveform of vibration signal from the eight engine states.

4.2. Feature Extraction Based on S Transform and 2DNMF. The S transform is utilized to convert the vibration signals from time domain to time-frequency domain, which can provide more discriminative information on engine working states. Figure 5 illustrates the time-frequency representations of eight engine states obtained by S transform.

For each sample of the engine, a 2048×4096 time-frequency matrix can be obtained based on the S transform. The size of the matrix is too large for processing. So the matrix is firstly segmented to 256×512 blocks, where every block consists of an 8×8 submatrix. The mean value of the submatrix is calculated to represent the block. In this way

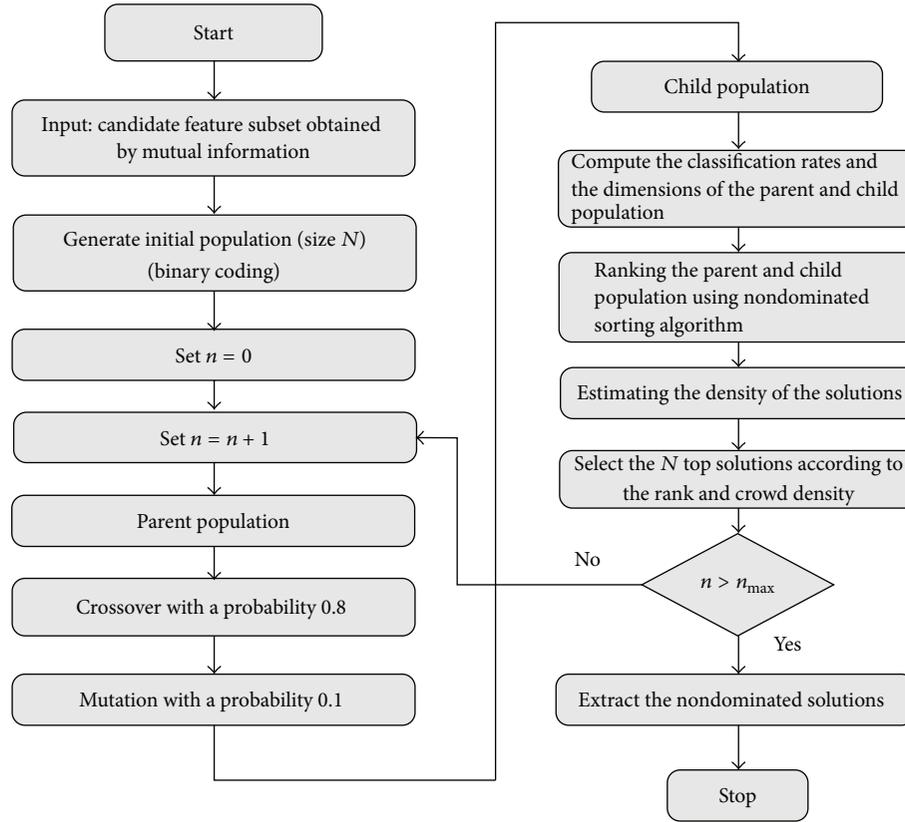


FIGURE 2: The schematic for applying the NSGA-II to the feature selection.

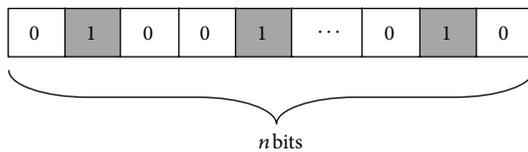


FIGURE 3: The binary coding system.

the time-frequency matrix reduced to 256×512 . The new matrix could provide enough information for classification of engine faults. But it is still not realistic to take all the elements of time-frequency distribution as features, where the dimension could be 131072. For this reason, it is very necessary to compress the feature dimensions from the time-frequency matrix. The 2DNMF, as described in Section 2.2, is used to characterize the time-frequency representations (TFRs) for engine fault diagnosis.

Based on the S transform, 320 TFRs in total can be obtained. Forty TFRs, five samples for each engine state, are selected as training samples to calculate the column basis matrix \mathbf{L} and row basis matrix \mathbf{R} for 2DNMF. The parameters d_1 and d_2 of the 2DNMF are both set to be 10. Figures 6 and 7 have shown the column basis matrix \mathbf{L} and row basis matrix \mathbf{R} obtained by 2DNMF for characterizing the TFRs of engine.

Then by mapping each TFR onto column basis matrix \mathbf{L} and row basis matrix \mathbf{R} based on (9), a 10×10 mapping coefficient matrix can be obtained. The feature matrices of

forty training TFRs are demonstrated in Figure 8, in which every row represents one engine state.

It can be found that the feature matrices obtained by 2DNMF can distinguish the eight engine operating states very effectively. The 100 elements of the feature matrix can be used as the original feature subset for engine fault diagnosis. The original feature subset is set to be F_{Original} .

4.3. Feature Selection Based on Mutual Information and NSGA-II. Although the original feature subset obtained by 2DNMF has shown to be effective in identifying the eight engine states, the dimension of the feature subset can be reduced and the classification accuracy can be improved via feature selection methods. In this subsection, the hybrid filter and wrapper scheme based on mutual information and NSGA-II is employed to find a more compact feature subset for engine fault diagnosis.

In the first stage, the filter method based on mutual information is used to obtain a candidate feature subset. Figures 9 and 10 show the relevance values of the 100 features and the redundancy values between the 100 features calculated by mutual information.

According to the relevance values and redundancy values calculated above, 100 sequential feature subsets $S_1 \subset S_2 \subset \dots \subset S_{100}$ can be obtained based on the mRMR criterion described in Section 3.2. The first 50 features, meaning half of the original feature subset, are selected as the candidate

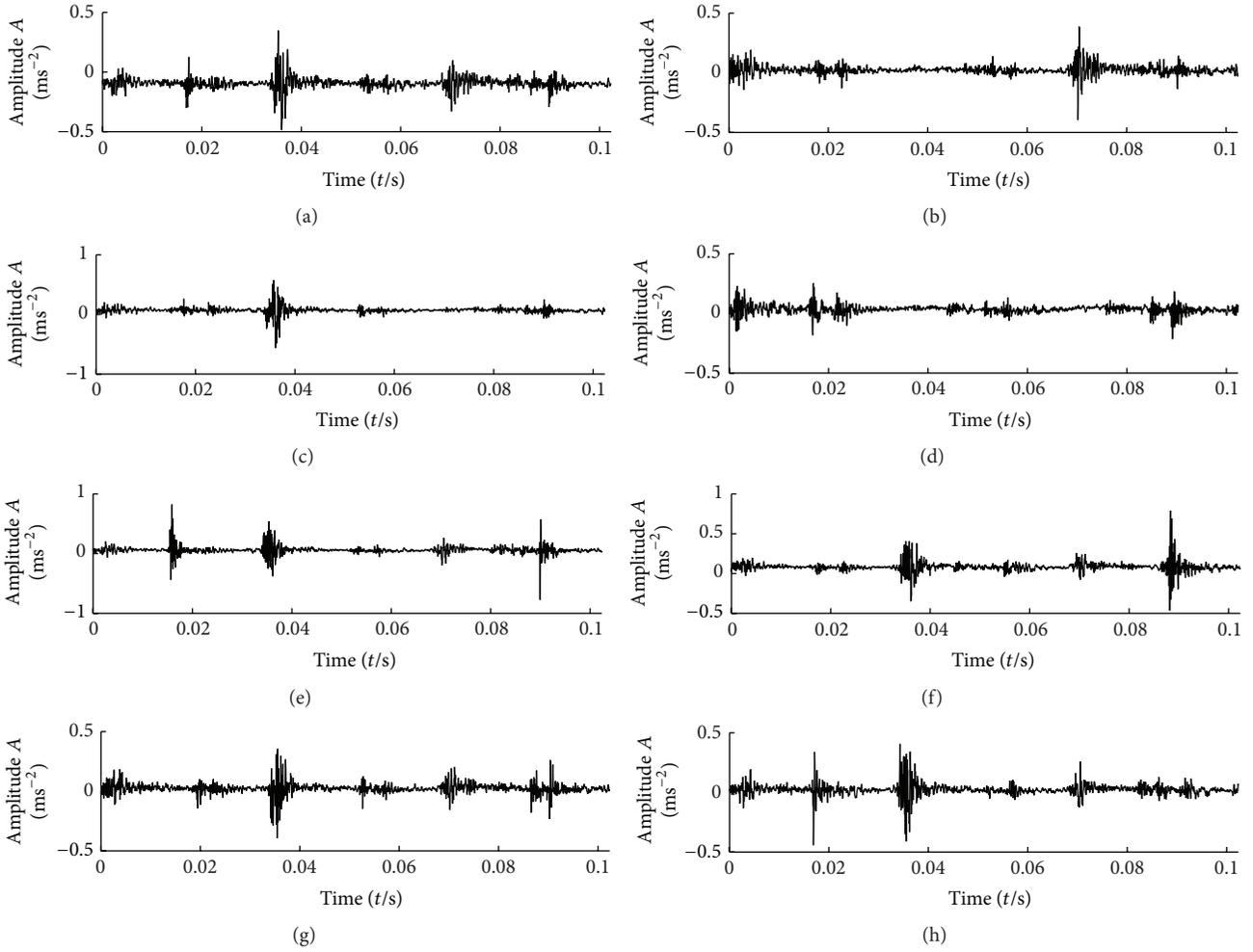


FIGURE 4: Vibration signals acquired from eight engine states: (a) normal state; (b) first cylinder misfire; (c) second cylinder misfire; (d) first and second cylinder misfire; (e) too large clearance of intake valve; (f) too large clearance of exhaust valve; (g) air leakage of the intake valve; (h) air leakage of the exhaust valve.

feature subset for wrapper methods in this work. This feature subset is denoted as F_{mRMR} .

In the second stage, the wrapper method based on NSGA-II is used to find a more compact feature subset based on the feature subset F_{mRMR} . An experiment is also conducted on applying the wrapper method directly on the original feature subset $F_{Original}$ for a comparison with the two-stage feature selection scheme.

Three classifiers, KNNC, NBC, and LS-SVM, as mentioned in Section 3.3, are employed to evaluate the presented two-stage feature selection scheme for engine fault diagnosis. The KNNC classifier is just selected to illustrate the wrapper method based on NSGA-II algorithm. For every chromosome created by NSGA-II, the dimension of features selected by this chromosome and the classification error rate based on KNNC classifier are regarded as the fitness functions. Other parameters of NSGA-II for feature selection are summarized in Table 2.

Figures 11 and 12 show the distributions of the solutions over the objective plane utilizing the NSGA-II and KNNC

TABLE 2: Parameters of NSGA-II for feature selection.

Parameters of NSGA-II	Parameter settings
Population size	200
Generation	60
Crossover probability	0.8
Mutation probability	0.01
Crossover method	Binary crossover
Mutation method	Binary mutation
Selection method	Tournament selection

based on the original feature subset $F_{Original}$ and the candidate feature set F_{mRMR} , respectively. In each figure, the solutions obtained by NSGA-II at generations 10, 20, 40, and 60 are shown. In the figures, all the solutions are marked as blue asterisks and the Pareto optimal solutions are marked with red circles. It can be observed that, with the increasing of generation, the Pareto fronts move toward the ideal

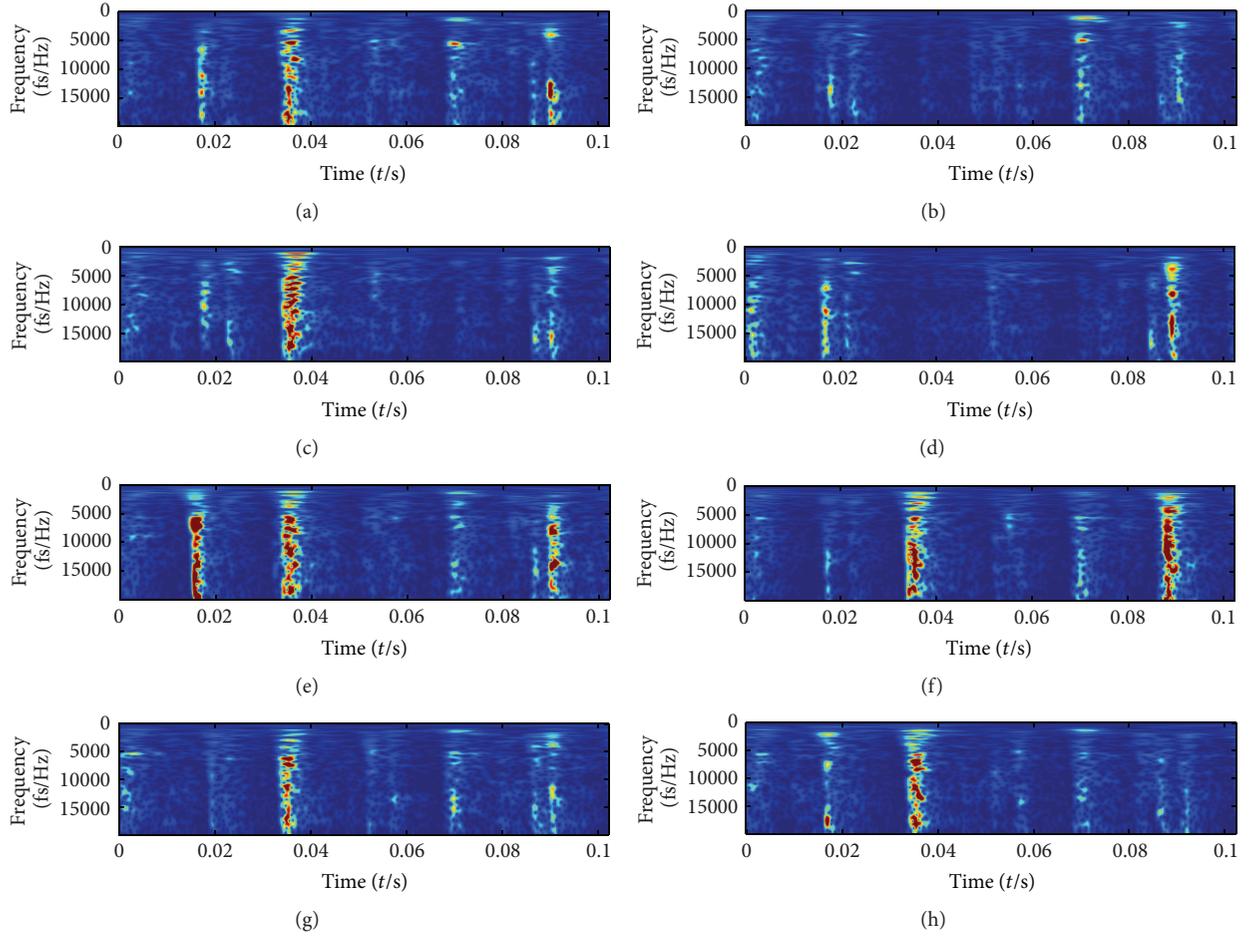


FIGURE 5: The S transform of the eight engine states: (a) normal state; (b) first cylinder misfire; (c) second cylinder misfire; (d) first and second cylinder misfire; (e) too large clearance of intake valve; (f) too large clearance of exhaust valve; (g) air leakage of the intake valve; (h) air leakage of the exhaust valve.

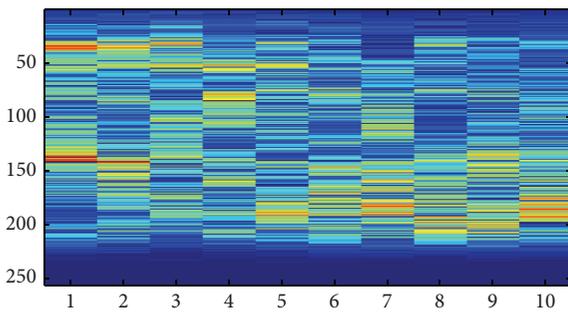


FIGURE 6: The column basis matrix obtained by 2DNMF.

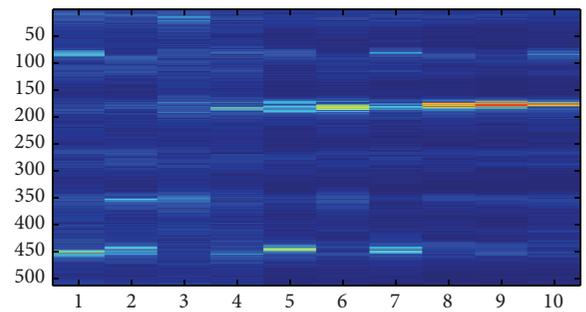


FIGURE 7: The row basis matrix obtained by 2DNMF.

solution, meaning the lower dimension of features and lower classification error rate, for the feature selection problem.

Furthermore, it can be obviously noted from Figures 11 and 12 that the hybrid feature selection scheme which conducts wrapper method on F_{mRMR} obtained better solutions than the wrapper method directly conducted on the original feature subset $F_{Original}$. The hybrid feature selection scheme

achieved a similar or lower classification error rate with much less features than the wrapper method.

4.4. Classification Performances of the Different Feature Subsets. In this section, the classification performance of four different feature subsets, the original feature subset ($F_{Original}$), the feature subset obtained by filter method based on mutual

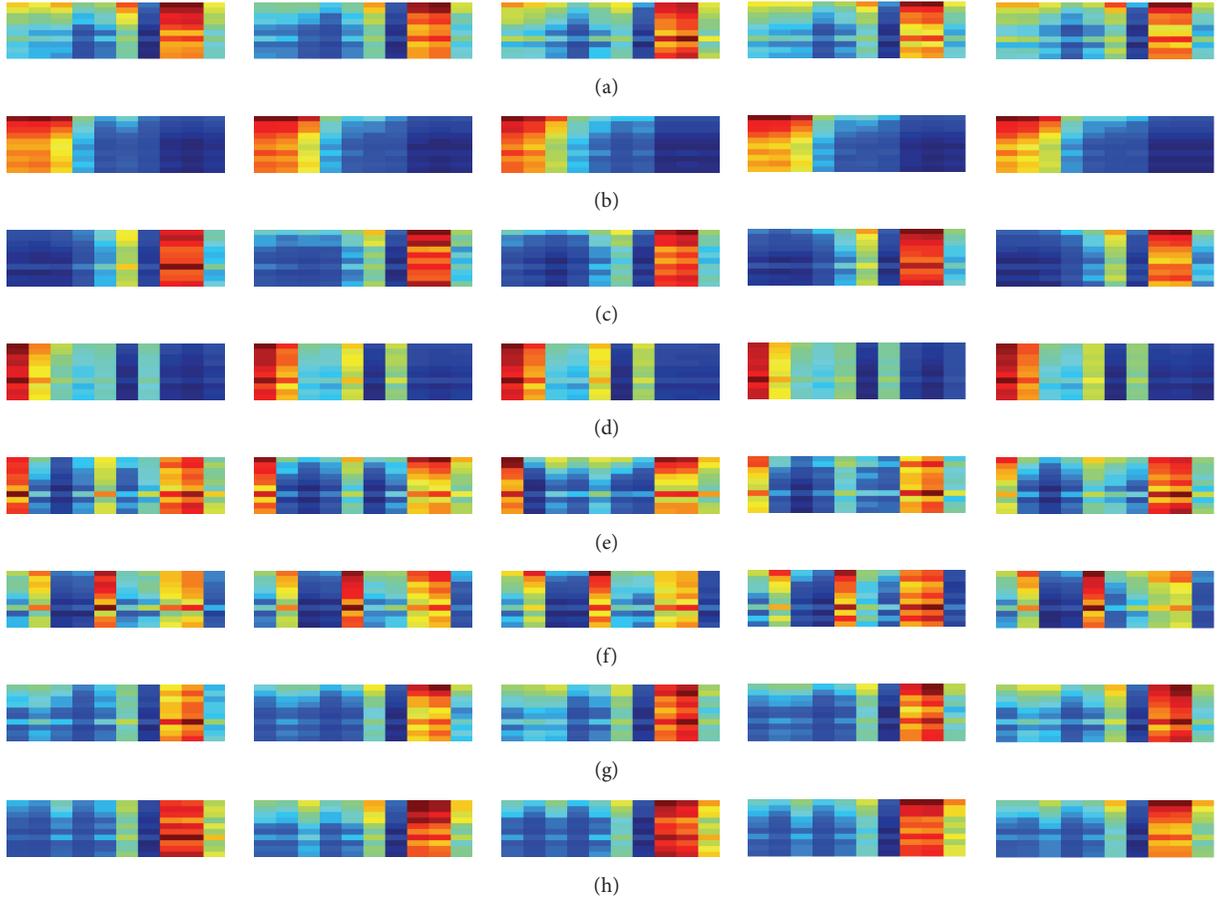


FIGURE 8: The feature matrices of eight engine states obtained by 2DNMF (every row represents one engine state and five samples of each engine state are displayed): (a) normal state; (b) first cylinder misfire; (c) second cylinder misfire; (d) first and second cylinder misfire; (e) too large clearance of intake valve; (f) too large clearance of exhaust valve; (g) air leakage of the intake valve; (h) air leakage of the exhaust valve.

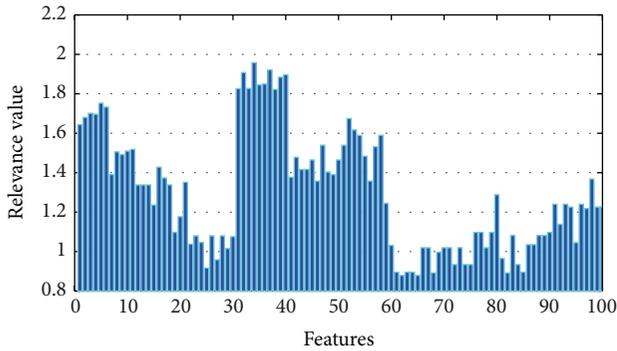


FIGURE 9: Relevance values of the features calculated by mutual information.

information (F_{mRMR}), the feature subset obtained by wrapper method based on NSGA-II (F_{NSGA}), and the feature subset obtained by the hybrid filter and wrapper method ($F_{\text{mRMR+NSGA}}$), is evaluated and compared. The computation time, feature subset dimensions, and the classification accuracy of different feature subsets based on three classifiers are shown in Table 3. All the experiments are conducted on a

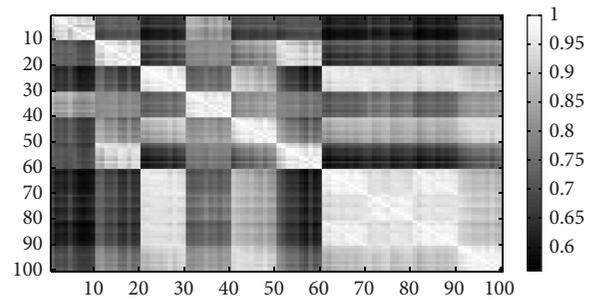


FIGURE 10: Redundancy values between the features calculated by mutual information.

personal computer with 2.93 GHz CPU and 512 M memory. The software used is MATLAB with version 7.1.

It can be found from Table 3 that the original feature subset F_{Original} , which consists of all the 100 features, obtained satisfactory classification performances by using all the three classifiers for engine fault diagnosis. It verifies the effectiveness of the presented feature extraction scheme based on S transform and 2DNMF.

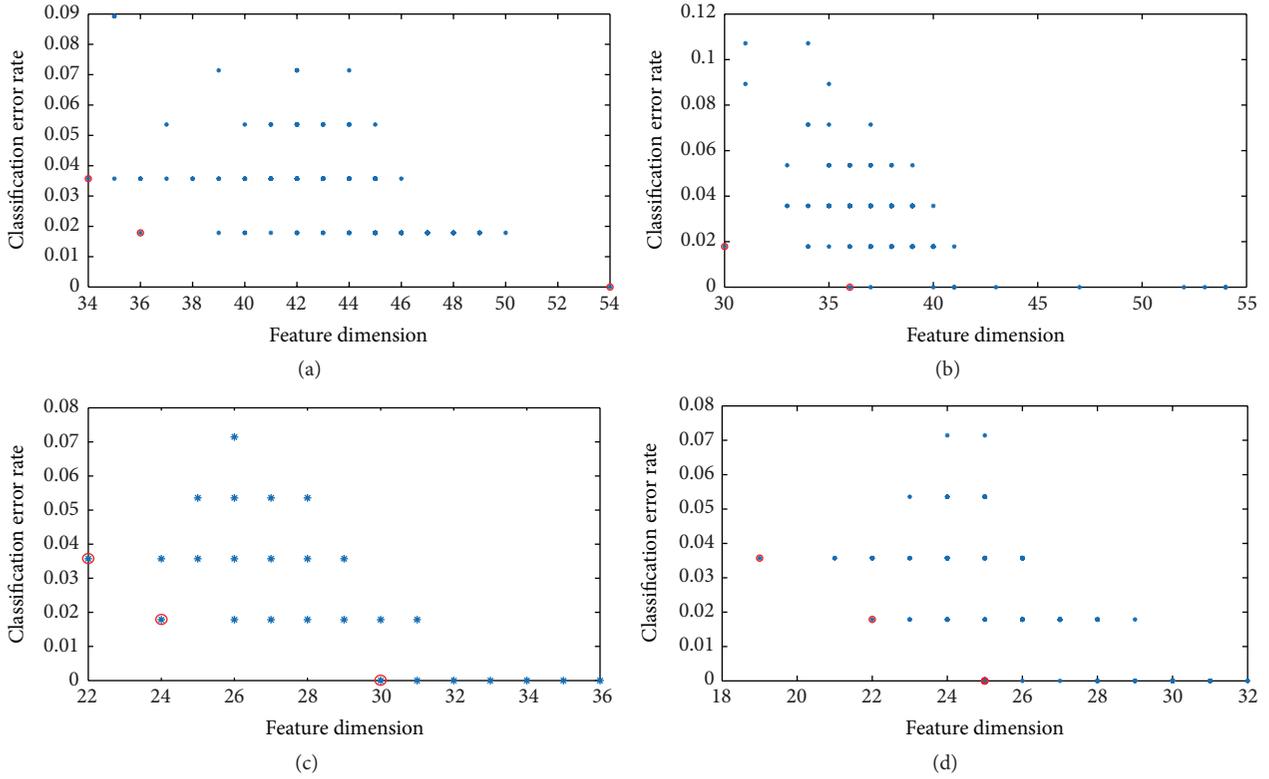


FIGURE 11: Distributions of the solutions at different generations over the objective plane utilizing the NSGA-II based on the original feature subset F_{Original} : (a) 10 generations; (b) 20 generations; (c) 40 generations; (d) 60 generations.

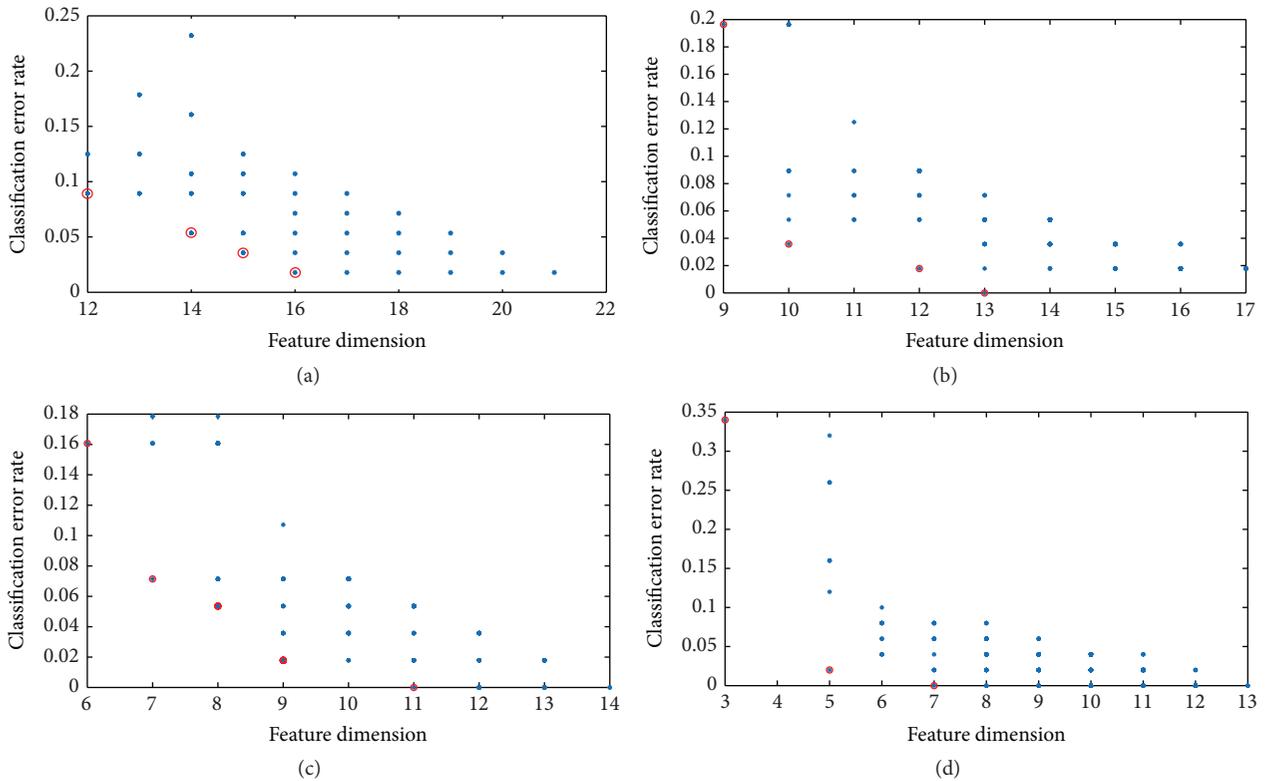


FIGURE 12: Distributions of the solutions at different generations over the objective plane utilizing the NSGA-II based on the candidate feature subset F_{mRMR} : (a) 10 generations; (b) 20 generations; (c) 40 generations; (d) 60 generations.

TABLE 3: Engine faults classification performances of four feature subsets F_{Original} , F_{mMRR} , F_{NSGA} , and $F_{\text{mRMR+NSGA}}$ based on three classifiers.

Classifiers	Feature subsets	Dimension of the feature subset	Computation time (s)	Classification accuracy (%)
KNNC	F_{Original}	100	—	92.5
	F_{mMRR}	40	2.54	96.25
	F_{NSGA}	25	102.3	100
	$F_{\text{mRMR+NSGA}}$	7	49.6	100
NBC	F_{Original}	100	—	87.5
	F_{mMRR}	40	2.54	88.75
	F_{NSGA}	24	148.25	96.25
	$F_{\text{mRMR+NSGA}}$	9	86.5	97.5
SVM	F_{Original}	100	—	96.25
	F_{mMRR}	40	2.54	97.5
	F_{NSGA}	27	496.4	100
	$F_{\text{mRMR+NSGA}}$	9	172.8	100

However, it also can be observed that the classification accuracies of F_{Original} are the lowest in the four feature subsets. It ascertains our assumption that there exist many irrelevant and redundant features, which will decrease the performances in the original feature subset. A feature selection procedure is indispensable before classification.

The performances of the feature subset F_{mRMR} showed to be better than F_{Original} . Otherwise, the dimension of F_{mRMR} is half of the original feature subset. However, the classification accuracies of F_{mRMR} are inferior to F_{NSGA} and $F_{\text{mRMR+NSGA}}$ obviously.

The feature subsets F_{NSGA} and $F_{\text{mRMR+NSGA}}$, which are obtained by wrapper method and hybrid filter and wrapper method, have demonstrated similar classification accuracies in our case. The $F_{\text{mRMR+NSGA}}$ achieved the highest classification rates by using KNNC and LS-SVM classifiers, while the F_{NSGA} gained the best performance by using NBC classifier. However, it can be found from Table 3 that the feature dimensions of $F_{\text{mRMR+NSGA}}$ are lower than F_{NSGA} . Moreover, the hybrid feature selection scheme required much less computation cost than the wrapper method. Therefore, it is very desirable to use hybrid filter and wrapper feature selection scheme to get a satisfactory refined feature subset for engine fault diagnosis.

It also can be found that very satisfactory classification performances can be achieved by the three classifiers with only 7, 9, and 9 features for engine fault diagnosis. It indicates that the presented feature extraction and selection scheme has provided a very effective and efficient approach for engine intelligent fault diagnosis.

5. Conclusion

This work has presented a new feature extraction and feature selection scheme for intelligent fault diagnosis of engine by utilizing S transform, 2DNMF, mutual information, and NSGA-II. Eight different engine operating states, which are simulated on an engine test rig, are employed to evaluate the effectiveness of the proposed methods.

In the phase of feature extraction, the S transform is firstly adopted to obtain time-frequency representations of vibration signals. Then the 2DNMF technique is applied to characterize the time-frequency representations. Experimental results revealed that the features extracted by S transform and 2DNMF obtain a very promising performance for identifying the eight engine states.

In the phase of feature selection, the hybrid filter and wrapper scheme based on mutual information and NSGA-II is employed to obtain a more compact feature subset and higher classification accuracy. And the performance of the hybrid method and the separate filter and wrapper methods is also compared. Experimental results have shown that the hybrid feature selection scheme achieved very promising performances with very few features for engine fault diagnosis. The faults classification accuracies of three different classifiers using the selected features by the presented scheme are consistently higher than those using original feature subset and feature subsets obtained by other feature selection methods. The dimension of the feature subset obtained by the hybrid feature selection scheme is lower than the filter or wrapper method. Furthermore, the computation cost of the hybrid feature selection is much less than the wrapper method.

This research demonstrates clearly that the presented feature extraction and feature selection scheme has great potential to be an effective and efficient tool for the fault diagnosis of engine and can be easily extended to be applied to other machineries.

Competing Interests

The authors declare that they have no competing interests.

References

- [1] Ş. Yildirim, S. Erkaya, I. Eski, and I. Uzmay, "Noise and vibration analysis of car engines using proposed neural network," *Journal of Vibration and Control*, vol. 15, no. 1, pp. 133–156, 2009.

- [2] J.-D. Wu, C.-K. Huang, Y.-W. Chang, and Y.-J. Shiao, "Fault diagnosis for internal combustion engines using intake manifold pressure and artificial neural network," *Expert Systems with Applications*, vol. 37, no. 2, pp. 949–958, 2010.
- [3] B. Li, R.-X. Hu, G.-Q. Ren, and J.-P. Fu, "Engine fault diagnosis based on a morphological neural network using a morphological filter as a preprocessor," *Proceedings of the Institution of Mechanical Engineers, Part D*, vol. 227, no. 4, pp. 490–505, 2013.
- [4] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert-Huang transform and support vector machine," *Applied Acoustics*, vol. 75, no. 1, pp. 1–9, 2014.
- [5] Y. Liu, J. H. Zhang, F. R. Bi, J. W. Lin, and W. P. Ma, "A fault diagnosis approach for diesel engine valve train based on improved ITD and SDAG-RVM," *Measurement Science and Technology*, vol. 26, no. 2, Article ID 025003, 2014.
- [6] P. K. Wong, J. H. Zhong, Z. X. Yang, and C. M. Vong, "Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis," *Neurocomputing*, vol. 174, pp. 331–343, 2016.
- [7] B. Li, P.-L. Zhang, S.-B. Liang, Y.-T. Zhang, and H.-B. Fan, "Feature extraction for engine fault diagnosis utilizing the generalized S-transform and non-negative tensor factorization," *Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science*, vol. 225, no. 8, pp. 1936–1949, 2011.
- [8] B. Li, S.-S. Mi, P.-Y. Liu, and Z.-J. Wang, "Classification of time-frequency representations using improved morphological pattern spectrum for engine fault diagnosis," *Journal of Sound and Vibration*, vol. 332, no. 13, pp. 3329–3337, 2013.
- [9] X. Wang, C. W. Liu, F. R. Bi, X. Y. Bi, and K. Shao, "Fault diagnosis of diesel engine based on adaptive wavelet packets and EEMD-fractal dimension," *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 581–597, 2013.
- [10] S. M. Jafari, H. Mehdigholi, and M. Behzad, "Valve fault diagnosis in internal combustion engines using acoustic emission and artificial neural network," *Shock and Vibration*, vol. 2014, Article ID 823514, 9 pages, 2014.
- [11] Y. Liu, J. H. Zhang, and L. Ma, "A fault diagnosis approach for diesel engines based on self-adaptive WVD, improved FCBF and PECOC-RVM," *Neurocomputing*, vol. 177, pp. 600–611, 2016.
- [12] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: the S transform," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 998–1001, 1996.
- [13] B. Li, P.-L. Zhang, D.-S. Liu, S.-S. Mi, G.-Q. Ren, and H. Tian, "Feature extraction for rolling element bearing fault diagnosis utilizing generalized S transform and two-dimensional non-negative matrix factorization," *Journal of Sound and Vibration*, vol. 330, no. 10, pp. 2388–2399, 2011.
- [14] G. G. Yen, "Wavelet packet feature extraction for vibration monitoring," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 3, pp. 650–667, 2000.
- [15] Q. Hu, Z. He, Z. Zhang, and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 688–705, 2007.
- [16] Y. Lei, Z. He, Y. Zi, and Q. Hu, "Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs," *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2280–2294, 2007.
- [17] V. Sugumaran, V. Muralidharan, and K. I. Ramachandran, "Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 930–942, 2007.
- [18] V. Sugumaran and K. I. Ramachandran, "Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing," *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2237–2247, 2007.
- [19] L. B. Jack and A. K. Nandi, "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms," *Mechanical Systems and Signal Processing*, vol. 16, no. 2-3, pp. 373–390, 2002.
- [20] Z. Chen, Y. He, F. Chu, and J. Huang, "Evolutionary strategy for classification problems and its application in fault diagnostics," *Engineering Applications of Artificial Intelligence*, vol. 16, no. 1, pp. 31–38, 2003.
- [21] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Engineering Applications of Artificial Intelligence*, vol. 16, no. 7-8, pp. 657–665, 2003.
- [22] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mechanical Systems and Signal Processing*, vol. 18, no. 3, pp. 625–644, 2004.
- [23] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and genetic algorithm for bearing fault detection," *Soft Computing*, vol. 10, no. 3, pp. 264–271, 2006.
- [24] B. Samanta and C. Nataraj, "Use of particle swarm optimization for machinery fault detection," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 2, pp. 308–316, 2009.
- [25] M. Danubianu, S. G. Pentiu, and D. M. Danubianu, "Data dimensionality reduction for data mining: a combined filter-wrapper framework," *International Journal Of Computers Communications & Control*, vol. 7, no. 5, pp. 824–831, 2012.
- [26] J. M. Cadenas, M. C. Garrido, and R. Martínez, "Feature subset selection Filter-Wrapper based on low quality data," *Expert Systems with Applications*, vol. 40, no. 16, pp. 6241–6252, 2013.
- [27] C. Peng, M. H. Wang, Y. Shen, H. Q. Feng, and A. Li, "Reconstruction and analysis of transcription factor-miRNA co-regulatory feed-forward loops in human cancers using filter-wrapper feature selection," *PLoS ONE*, vol. 8, no. 10, Article ID e78197, 2013.
- [28] Z. Y. Hu, Y. K. Bao, T. Xiong, and R. Chiong, "Hybrid filter-wrapper feature selection for short-term load forecasting," *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 17–27, 2015.
- [29] B. Li, P. Zhang, G. Ren, and Z. Xing, "A two stage feature selection method for gear fault diagnosis using relief and GA-wrapper," in *Proceedings of the International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '09)*, pp. 578–581, Zhangjiajie, China, April 2009.
- [30] B. Li, P.-L. Zhang, H. Tian, S.-S. Mi, D.-S. Liu, and G.-Q. Ren, "A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10000–10009, 2011.
- [31] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [32] M. Sebban and R. Nock, "A hybrid filter/wrapper approach of feature selection using information theory," *Pattern Recognition*, vol. 35, no. 4, pp. 835–846, 2002.

- [33] Ö. Uncu and I. B. Türkşen, "A novel feature selection approach: combining feature wrappers and filters," *Information Sciences*, vol. 177, no. 2, pp. 449–466, 2007.
- [34] Z. Zhu, Y.-S. Ong, and M. Dash, "Wrapper-filter feature selection algorithm using a memetic framework," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 1, pp. 70–76, 2007.
- [35] M.-C. Lee, "Using support vector machine with a hybrid feature selection method to the stock trend prediction," *Expert Systems with Applications*, vol. 36, no. 8, pp. 10896–10904, 2009.
- [36] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [37] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.
- [38] N. Srinivas and K. Deb, "Multiobjective optimization using nondominated sorting in genetic algorithms," *Evolutionary Computation*, vol. 2, no. 3, pp. 221–248, 1994.
- [39] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [40] L. B. Jack and A. K. Nandi, "Genetic algorithms for feature selection in machine condition monitoring with vibration signals," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 147, no. 3, pp. 205–212, 2000.
- [41] P. J. Grother, G. T. Candela, and J. L. Blue, "Fast implementations of nearest neighbor classifiers," *Pattern Recognition*, vol. 30, no. 3, pp. 459–465, 1997.
- [42] R. R. Yager, "An extension of the naive Bayesian classifier," *Information Sciences*, vol. 176, no. 5, pp. 577–588, 2006.
- [43] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [44] P. J. R. P. W. Duin, P. Paclik, E. Pekalska, D. de Ridder, D. M. J. Tax, and S. Verzakov, *PRTools4.1, A Matlab Toolbox for Pattern Recognition*, Delft University of Technology, Delft, The Netherlands, 2007.
- [45] LS-SVMlab1.5, <http://www.esat.kuleuven.ac.be/sista/lssvmlab/>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

