

Received January 31, 2020, accepted February 15, 2020, date of publication February 19, 2020, date of current version February 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975113

Engine Working State Recognition Based on Optimized Variational Mode Decomposition and Expectation Maximization Algorithm

XIAOBO BI^{1,2}, JIANGSHENG LIN¹, FENGRONG BI¹, XIN LI¹, DAIJIE TANG¹,
YOUXI WU^{2,3}, (Member, IEEE), XIAO YANG¹, AND PENGFEI SHEN¹

¹State Key Laboratory of Engines, Tianjin University, Tianjin 300072, China

²School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China

³Hebei Province Key Laboratory of Big Data Calculation, Hebei University of Technology, Tianjin 300401, China

Corresponding author: Fengrong Bi (fr_bi@tju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Project 51705357, and in part by the Tianjin Sci-Tech Project under Grant 2169901.

ABSTRACT Increasingly energy and environmental crises put forward higher request on diesel engine. It promotes the development of diesel engine, while the complexity of structure is much higher, which leads to higher probability of faults. In order to recognize the states of engine in harsh environments effectively, variational mode decomposition (VMD) and expectation maximization (EM) are introduced into this paper to analyze multi-channel vibration signals. To select the decomposition level of VMD adaptively, a novel power spectrum segmentation based on scale-space representation is proposed for the optimization of VMD and results show this approach can discriminate different frequency components in high noise circumstance accurately and efficiently. To improve the adaptability and accuracy of EM, a feature selection approach based on genetic algorithm (GA) is introduced to preprocess original data and a cross validation method is used for selecting cluster number adaptively. Combined with these approaches, a diesel engine state recognition scheme based on multi-channel vibration signals using optimized VMD and EM is proposed. Compared with existing method, this scheme shows great advantages in accuracy and efficiency, and could be applied in actual engineering.

INDEX TERMS Diesel engine, vibration, pattern recognition, variational mode decomposition (VMD), expectation maximization (EM).

I. INTRODUCTION

As a frequently used power source, diesel engines are being widely employed in industry, agriculture and some special industries because of the advantages of low fuel consumption and large output torque [1]. With increasingly energy and environmental crises, a large number of laws and regulations for diesel engine have been enacted in many countries, which greatly promotes the development of new technologies such as turbo charging, high-pressure common-rail fuel system, electronically controlled fuel injection system, variable valve timing (VVT) and exhaust gas recirculation (EGR) and so on [2]. While the new technologies improve the performance of diesel engine significantly, complex structure increases the

probability of failure. As one of the main parts in mechanical system, engine may lead serious security incidents under failure, so fault detection is an important research direction at present.

In order to ensure normal operation, fault detection system should recognize failures in early stage and issue a warning timely. So engine fault detection strategy develops from regular and disassembling diagnosis to real-time and non-disassembly diagnosis [3]. When developing the strategy, direct signal, such as rotational speed, temperature and pressure, is a traditional choice. For example, Xu *et al* used shaft instantaneous angular speed to extract characteristic parameters in order to detect misfire [4]. Although the direct signal is simple, it can hardly deal with other common faults such as mechanical failure and wear. To detecting various faults, some real-time acquiring indirect signal, such as noise

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wang¹.

and vibration, is being a promising method, in which the vibration signal is a research focus because of easy measurement, low cost and strong robustness. For example, Jing *et al* analyzed vibration signals from cylinder head and detected valve clearance fault by it [5], Ftoutou *et al* used vibration signals to make research on fuel injection faults [6].

Owing to complex structure, so many vibration sources exist in engine that there is much noise in vibration signals. A number of signal processing algorithms have been researched to overcome this problem such as wavelet transform and empirical mode decomposition (EMD). Moosavian *et al* analyzed the performances of different mother wavelets and used wavelet transform to denoise vibration signals [7], Ma *et al* mixed wavelet transform and EMD to analyze vibration signals in order to diagnose abnormal combustion in engine [8], Li *et al* detected abnormal clearance between contacting components using EMD [9]. However, the wavelet transform used for signal decomposition is generally binary discrete wavelet transform (DWT), which just decomposes signals in Fourier spectrum mechanically. Moreover, wavelet basis has a great influence on decomposition while there is no definite conclusion on how to choose it. As an adaptable method, the EMD is built based on recursive model, which results in low robustness and mode mixing [10], [11]. In 2013, Gilles proposed empirical wavelet transform (EWT), which can build adaptable wavelet filter to extract components [12]. For the problem that meaningful modes should be pre-detected to provide necessary parameters for the wavelet filter, Gilles *et al* proposed a parameterless scale-space approach to overcome it [13], which improved the EWT and promoted its application in faults detection such as Wang *et al* used EWT analyzed the fault features of industrial bearing through vibration signals [14]. However, the problem of wavelet basis is still unresolved. Besides, the process of building wavelet filter is complex and the performance of filter is no guarantee of the best. In 2014, Dragomiretskiy *et al* proposed variational mode decomposition (VMD) based on variational model, which is an essentially adaptable Wiener filter bank and it is equating to the best filter under certain conditions [15]. The VMD has been used in faults detection for rotating machinery and showed great potential [16]. However, decomposition level in VMD should be determined manually, which increases the complexity and uncertainty. Fortunately, the parameterless scale-space approach also has ability to provide parameters for VMD. So the method is optimized to combine with VMD in order to build a better filter.

Although a lot of successes have been achieved in the diesel engine condition recognition and fault diagnosis based on time-frequency analysis, feature extraction and pattern recognition of single channel vibration signal, there are still some problems need solving. Firstly, in some harsh environments, unpredictable signal interference may lead to the misinterpretation of recognition rate. Secondly, every feature parameter has different sensitivity to different working states, which results in some features or feature

combinations may not identify the working state effectively in certain case [17], [18]. For these reasons, multi-sensor information fusion technology is gradually attached the weight of researchers. Multi-sensors information resources can be fully and effectively utilized through data fusion method to gain the greatest resources of system under test in different view angles [19]. However, the increase of features has a negative impact, such as information redundancy and complex data processing on classification, and even reduce the recognition rate [20]. Accordingly, it is necessary to reduce the feature dimension by removing redundant and irrelevant features in order to improve classification efficiency and accuracy. Feature selection is a challenging task, especially for hundreds or thousands of features sub sets, in which genetic algorithm (GA) is a classic method to improve the searching efficiency. Stefano *et al* proposed an improved filter-based GA to search feature subsets with high discriminative power [21]. Mohammadi *et al* presented a methodology which can search and select a set of closure relationships for given experimental based on GA to minimize the error between measured and predicted pressure gradient [22]. Gangavarapu *et al*, used GA to optimize the subspace ensembling process of high-dimensional biomedical datasets feature subspaces [23]. The performance of GA highly depends on the selection of fitness function. However, various subsets of dataset may give best results with a different function [24].

Feature selection is a preprocessing technique for classification algorithm [25], [26]. X-means clustering algorithm is an extending K-means with efficient estimation of the number of clusters. It goes into action after each run of K-means, making local decision about which subset of current centroids should split themselves in order to better fit the data. The splitting decision is done by the Bayesian information criterion (BIC) [27], and the numbers of clusters are computed dynamically within preset lower and upper bound [28]. X-means algorithm has a good effect on processing low-dimensional dataset, while its performance on high-dimensional dataset is unstable. Expectation maximization (EM) algorithm based on Gaussian mixture model has been being widely used due to its stability and simplicity [29]. Zhao *et al* proposed an improved EM algorithm for fault detection of air conditioning system [30]. Chen *et al* put forward an adaptive Gaussian mixture model to deal with the dynamic working process of rotating turbine engine disk, which improved the adaptability of fault detection model to turbine engine disk [31]. Tapanan *et al* evaluated the performance of various clustering techniques and results showed that the EM clustering algorithm is the most effective and robust method for unlabeled data classification [32]. However, the conventional EM algorithm cannot choose the number of clusters adaptively and need to specify the number of clusters as an input parameter.

The paper is organized as follows: section 1 introduces the research background and significance, section 2 gives algorithms details, the experiment is described in section 3, the optimized VMD is proposed in section 4, analysis

and contrast of experiment result by a novel approach based on EM are shown in section 5 and conclusion is given in section 6.

II. ALGORITHM THEORIES

A. VMD THEORIES

The goal of VMD is to decompose input signal f into several modes, u_k , i.e. intrinsic mode functions (IMFs), which are supposed to have specific sparsity properties and limited bandwidths. The IMFs are also assumed to be mostly compact around a center frequency, ω_k [15].

Firstly, the u_k is processed by Hilbert transform to get unilateral frequency spectrum:

$$Hu_k = \frac{1}{\pi} p.v. \int_{\mathbb{R}} \frac{u_k(v)}{t-v} dv = \left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \quad (1)$$

where $p.v$ is Cauchy principal value, δ is Dirac distribution, $*$ represents convolution and $j^2 = -1$. $k \in [1, 2, \dots, K]$, where K is the decomposition level in VMD, i.e. the number of IMFs.

Next, an estimated center frequency, $e^{j\omega_k t}$, is mixed to shift the unilateral frequency spectrum into baseband:

$$B = \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \quad (2)$$

Bandwidth can be computed by squared L^2 -norm of the gradient. Meanwhile, all of the IMFs can restructure the input signal f . Based on that, a constrained variational model can be obtained:

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \\ \text{s.t.} & \sum_k u_k = f \end{aligned} \quad (3)$$

where $\{u_k\} := \{u_1, u_2, \dots, u_K\}$ and $\{\omega_k\} := \{\omega_1, \omega_2, \dots, \omega_K\}$ are all IMFs and their center frequencies, respectively. $\sum := \sum_{k=1}^K$ is the summation of all IMFs.

In order to solve the constrained variational model, Lagrangian multipliers, λ , and quadratic penalty α , are introduced to shift the model into unconstrained. λ and α are two classic ways to solve variational problem. The α is used to enforce constraints strictly and it can be ignored at low requirements for constraint. The α is used to balance the data-fidelity constraint especially in high Gaussian noise condition and it is inversely proportional to the noise level in data based on Bayesian prior. The unconstrained variational model is:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) &= \\ &= \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (4)$$

Based on Parseval / Plancherel Fourier isometry, it can be solved as:

$$\widehat{u}_k^{n+1}(\omega) = \frac{\widehat{f}(\omega) - \sum_{i \neq k} \widehat{u}_i(\omega) + \frac{\widehat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (5)$$

$$\omega_k^{n+1}(\omega) = \frac{\int_0^\infty \omega \left| \widehat{u}_i(\omega) \right|^2 d\omega}{\int_0^\infty \left| \widehat{u}_i(\omega) \right|^2 d\omega} \quad (6)$$

$$\widehat{\lambda}^{n+1}(\omega) = \widehat{\lambda}^n(\omega) + \tau \left(\widehat{f}(\omega) - \sum_k \widehat{u}_k^{n+1}(\omega) \right) \quad (7)$$

where n is the number of iterations, τ is time-step of the dual ascent and it is set as 0, i.e. taking no account of λ in this paper.

Based on that, IMFs can be obtained by alternate direction method of multipliers (ADMM). The process of VMD is:

- ① Initialize $\{\widehat{u}_k^1\}$, $\{\widehat{\omega}_k^1\}$, $\widehat{\lambda}^1$ and n .
- ② Compute u_k , ω_k and λ based on equation(5), (6) and (7) through ADMM.
- ③ During the iteration in step ②, results can be obtained when stopping condition $\sum_k \left\| \widehat{u}_k^{n+1} - \widehat{u}_k^n \right\| / \left\| \widehat{u}_k^n \right\| < \varepsilon$ is met, where $\varepsilon = 10^{-7}$ in general.

B. GENETIC ALGORITHM BASED FEATURES SELECTION

GA is a method to solve optimization inspired by biological processes of mutation, natural selection and genetic crossover, which is also a powerful feature selection tool, especially for feature set with large dimensions

The basic genetic operators of GA are selection, crossover, and mutation [33].

1) SELECTION

Selection is based on individual fitness and influences its ability to reproduce into the next generation [34]. The probability of selecting individual h_i is determined by (8):

$$p(h_i) = \frac{F(h_i)}{\sum_{i=1}^p F(h_i)} \quad (8)$$

where $F(h_i)$ is the fitness value of h_i . The probability that an individual will be selected is proportional to its own fitness and is inversely proportional to the fitness of the other competing hypothesis in the current population.

2) CROSSOVER

Select two chromosomes than have high fitness values from the current population, exchange some bits of them and copy them into the new population. The location of these bits are random.

3) MUTATION

Select one chromosome that has a high fitness value from the current population, alter some bits of it and copy it into the new population [35].

C. EXPECTATION-MAXIMIZATION ALGORITHM FOR GAUSSIAN MIXTURES

Given a Gaussian mixture model, the goal is to maximize the likelihood function with respect to the parameters comprising the means and covariances of the components and the mixing coefficients. The main processes are shown as follows.

① Initialize means μ_k , covariances Σ_k and mixing coefficients π_k , and evaluate the initial value of log likelihood.

② **E step:** Evaluate responsibilities based on current parameter values.

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)} \quad (9)$$

③ **M step:** Re-estimate the parameters based on current responsibilities

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n \quad (10)$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T \quad (11)$$

$$\pi_k^{new} = \frac{N_k}{N} \quad (12)$$

where

$$N_k = \sum_{n=1}^N \gamma(z_{nk}) \quad (13)$$

④ Evaluate the log likelihood:

$$\ln p(X | \mu, \Sigma, \pi) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k) \right\} \quad (14)$$

Check for the convergence of parameters or the likelihood, and return to the Step ② if the convergence criterion is not satisfied [36].

EM assigns probability distribution belonging to certain cluster for each instance.

III. EXPERIMENT

In order to research the approach of engine faults detection, author team built experiment system and designed experiment project. The experiment system consists of vibration testing and engine bench, as shown in Fig. 1. The engine bench system includes experimental engine and electrical dynamometer. The electrical dynamometer is manufactured by AVL List GmbH. The experimental engine is a turbo charged in-line 6-cylinder diesel engine and its parameters is listed in Table 1.

The engine faults were designed and simulated on this bench test. Statistically, injection mechanism (27.0%), water leakage (17.3%) and valve and its seat are the most frequent faults in diesel engine [3]. Among these, water leakage generally leads to high water temperature, which can be easily

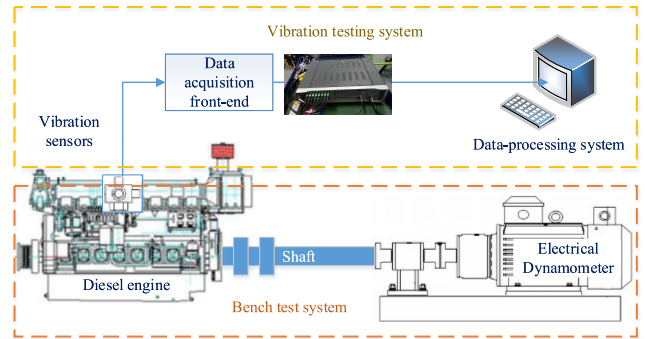


FIGURE 1. Experiment system.

TABLE 1. Parameters of the experimental engine.

Items	Parameters
Displacement	7.14L
Firing sequences	1-5-3-6-2-4
Rated power	220kw@2300r/min
Maximum torque	1250Nm@1200r/min-1600r/min

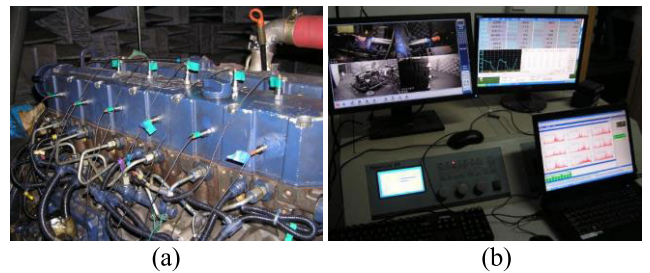


FIGURE 2. Engine bench test. (a) Bench test. (b) Vibration testing system.

identified by instrumentations. So, the rest two faults are the main research objects in this paper. The inappropriate valve clearance is mainly manifested in valve and its seat fault caused by assembly error or abrasion, for which big and small valve clearances are researched. The injection mechanism fault is complex, and it is divided into injection timing and rail pressure in this paper.

Vibration testing system includes acceleration sensors, data acquisition and processing system. During engine operation, cylinder explosion pressure and valve bounce have a direct impact on cylinder head, so the acceleration sensors are placed on cylinder head cover to collected vibration signals with high signal-to-noise ratio (SNR) and rich information, as shown in Fig. 2. Main parameters of testing equipment are listed in Table 2.

Considering that the characteristic frequency of engine fault is below 10 kHz, sampling frequency is set as $f_s = 25.6\text{kHz}$. The experiment simulates the failure of valve gap, different rail pressure parameters and different injection timing. Speed conditions are selected as 1600r/min and 2100r/min, load condition and other operating parameters setting are shown in Table 3.

TABLE 2. Main parameters of testing equipment.

Testing equipment	Model
Data acquisition front-end	LMS SCM05
Vibration sensor	PCB 621B40
Laptop computer	ThinkPad T530

TABLE 3. Operating parameters setting.

Injection timing (50% load)	Rail pressure (100% load)	Valve clearance (100% load)
3°C A	1000bar	Increase (0.5-0.6)
5°C A	1100bar	Normal (0.4-0.5)
7.5°C A (N)	1200bar	Reduction (0.3-0.4)
10°C A	1340bar(N)	

Where the valve clearance ‘(0.4, 0.5)’ means that the inlet valve clearance is 0.4 mm, the exhaust valve clearance is 0.5 mm, and so does the rest; tag ‘N’ means normal operating state parameter.

IV. SIGNAL DECOMPOSITION APPROACH BASED ON OPTIMIZED VMD

The decomposition level K in VMD should be selected manually, which increases complexity and subjectivity. An unsuitable K could result in over or under decomposition and influences the accuracy of subsequent analysis. To solve this problem, spectrum segmentation is introduced to optimized VMD.

A. OPTIMIZED VMD BASED ON FOURIER SPECTRUM SEGMENTATION

Gilles *et al* proposed a spectrum segmentation approach to detect meaningful modes for EWT [13]. VMD is a better filter and the selection of decomposition level K is essentially detecting meaningful modes. The spectrum segmentation approach is optimized to build VMD filter in this section.

Suppose the Fourier spectrum function of f is $F(x)$. $g(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$ is a kernel function, where t is scale parameter. A scale-space representation of $F(x)$ is

$$L(x, t) = g(x; t) * F(x) \quad (15)$$

where $*$ represents convolution.

The number of minima with respect to x of $L(x, t)$ is a decreasing function of the scale parameter t , and no new minima appear as t increases. Every minima corresponds to a curve in scale-space, and the length of curve is L_i . When decomposing signal with VMD, selecting decomposition level can be transformed into finding boundaries delimiting consecutive modes that is finding local minima in Fourier spectrum. Based on scale-space representation, this problem is equivalent to find curves which meet a certain threshold.

The threshold is selected by the principle of histogram. Firstly, sort the curves of local minima in ascending order by

their length and obtain $L_{sort} = \{L_1, L_2, \dots, L_M\}$, where M is the number of local minima. Next, suppose $\Delta = \frac{L_M - L_1}{L_M}$ and $L_{threshold} = L_M - n\Delta$, where n is a positive integer less than L_M . Finally, decrease the n one by one until the proportion of $L_i \in [L_{threshold}, L_M]$ to L_{sort} is just less than $1/M$, and the $L_{threshold}$ is taken as final threshold. Suppose the number of boundaries is N , the decomposition level of VMD is $K = N + 1$ through considering of frequency range (i.e. $[0, fs/2]$).

Unlike EWT, VMD needs center frequencies, instead of the boundaries, to build filter. The frequency at peak value of every segment in Fourier spectrum is taken as the center frequency, and the K center frequencies can be obtained: $\{\omega_1, \omega_2, \dots, \omega_K\}$. So the adjustment added to step ① in VMD algorithm is:

Analyze input signal by spectrum segmentation approach, set the decomposition level K , initialize the center frequency $\{\omega_k^1\}$ as $\{\omega_1, \omega_2, \dots, \omega_K\}$.

The rest processes remain unchanged and the desired VMD result can be obtained.

B. VERIFICATION WITH SIMULATED SIGNAL

To verify the advantages of the optimized VMD, a simulated signal is selected to be decomposed by the VMD and the EWT. The simulated signal is shown as equation (16).

$$\begin{cases} s_1 = \begin{cases} 150e^{(-700t)} \sin(2\pi f_1 t) & \text{if } t \in [4.0 \times 10^{-4}, 1.9 \times 10^{-3}] \\ 130e^{(-900t)} \sin(2\pi f_1 t) & \text{if } t \in [2.7 \times 10^{-3}, 4.2 \times 10^{-3}] \\ 110e^{(-1000t)} \sin(2\pi f_1 t) & \text{if } t \in [5.1 \times 10^{-4}, 6.5 \times 10^{-3}] \\ 100e^{(-1200t)} \sin(2\pi f_1 t) & \text{if } t \in [7.4 \times 10^{-4}, 8.9 \times 10^{-3}] \\ 0 & \text{else} \end{cases} \\ s_2 = 20 \sin(2\pi f_2 t) \\ s_3 = 30 \sin(2\pi f_3 t) \\ s_4 = \eta \\ s = s_1 + s_2 + s_3 + s_4 \end{cases} \quad (16)$$

where $t \in [0, 0.01]$, $f_1 = 6500$, $f_2 = 3000$, $f_3 = 10000$, η , i.e. s_4 , represents Gauss white noise of 25dBw. Sampling frequency is set as 25.6 kHz.

Based on that, the simulated signal and its Fourier spectrum are plotted as Fig. 3, in which the s_1 , s_2 and s_3 simulate periodic impact component, low frequency sinusoidal component and high frequency sinusoidal component, respectively.

Based on the scale-space segmentation approach, the Fourier spectrum is segmented as Fig. 4. Special to note is that the amplitude of Fourier spectrum is absolute value of single-sided fast Fourier transform (FFT) instead of the real value. The real value is calculated for the absolute value by linear transformation, which means analyzing the absolute value directly is a more efficient way.

As shown in Fig. 4, the Fourier spectrum segmentation approach has ability to divide components with different frequencies. Based on the segmentation result, VMD and EWT

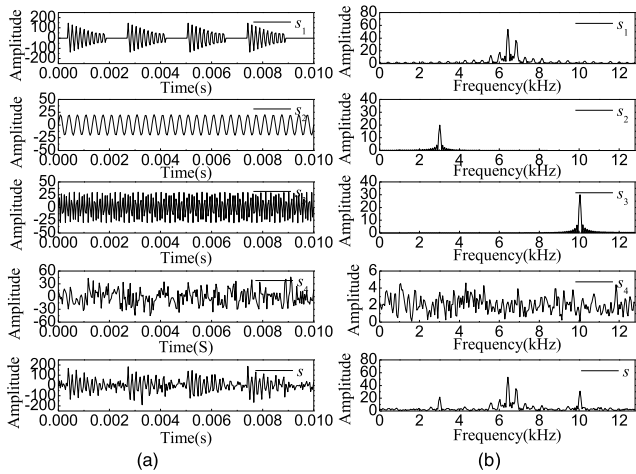


FIGURE 3. Simulated signal. (a) Time domain. (b) Frequency domain.

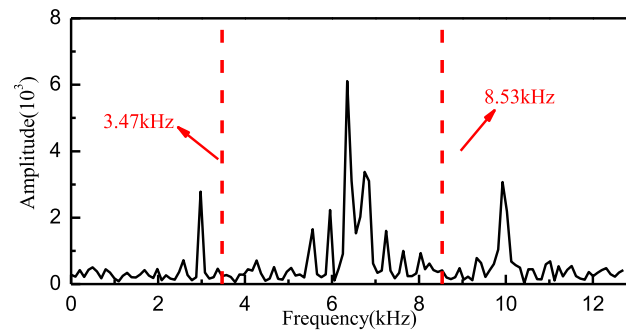


FIGURE 4. Fourier spectrum segmentation of simulated signal.

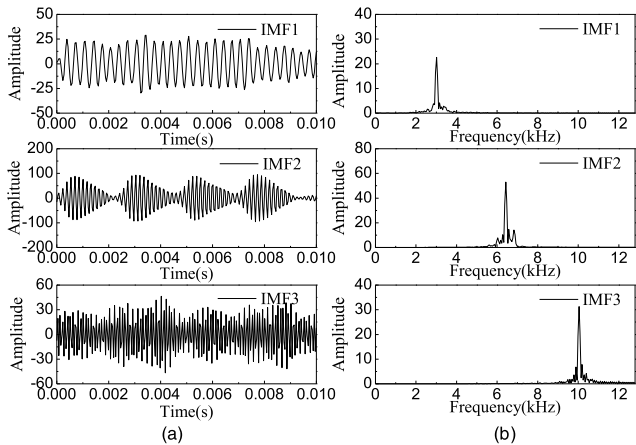


FIGURE 5. Decomposition result of simulated signal based on VMD. (a) Time domain. (b) Frequency domain.

filters can be built and their decomposition results are shown as Fig. 5 and Fig. 6, respectively. In the VMD decomposition, the quadratic penalty is selected as $\alpha = 2200$ by data analysis.

As shown in Fig. 5 and Fig. 6, VMD obtains better result compared with EWT. For the result of VMD, the IMFs are near to desired narrow-bandwidth original signals. For the results of EWT, IMF1 and IMF2 contain much noise

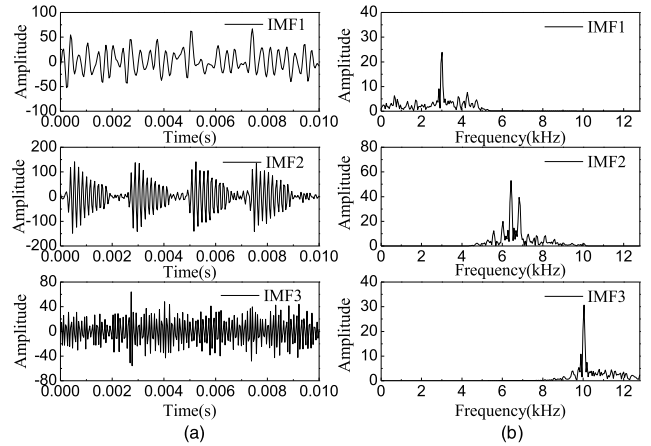


FIGURE 6. Decomposition result of simulated signal based on EWT. (a) Time domain. (b) Frequency domain.

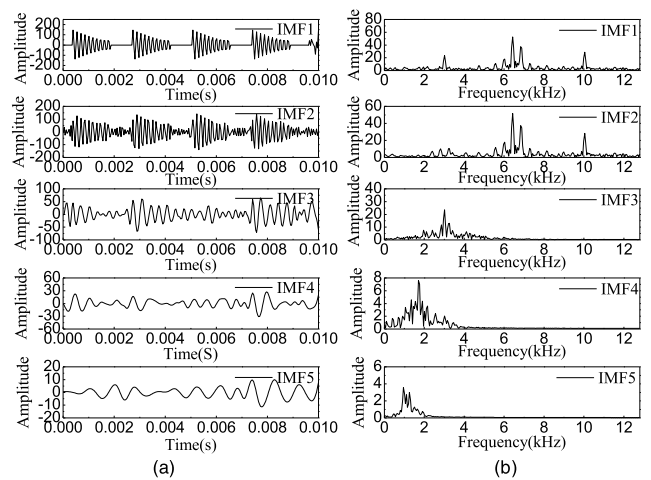


FIGURE 7. Decomposition result of EEMD. (a) Time domain. (b) Frequency domain.

compared with the s_2 and s_3 respectively, which are sinusoidal signals with single frequency (as shown in Fig. 3). To explain the advantages of VMD, traditional decomposing algorithms, ensemble empirical mode decomposition (EEMD) and local mean decomposition (LMD) are employed to decompose the simulated signal and the results are shown in Fig. 7 and Fig. 8. There are 9 IMFs in the result of EEMD and 7 IMFs in the results of LMD, and the first 5 IMFs in each result are shown for the rest ones are obvious illusive components.

As shown in Fig. 7 and Fig. 8, neither EEMD nor LMD could decompose high frequency components completely, which means s_1 and s_3 cannot be distinguished. The bandwidths of IMF1 and IMF2 are so large that much noise exists in them, which effects the recognition of faults badly. The reason for this shortcoming is that EEMD and LMD are built based on recursive model and have low robustness.

Because of low robustness, EWT, EEMD and LMD have no advantages in decomposing signals with complex frequency and high noise, such as engine vibration signals.

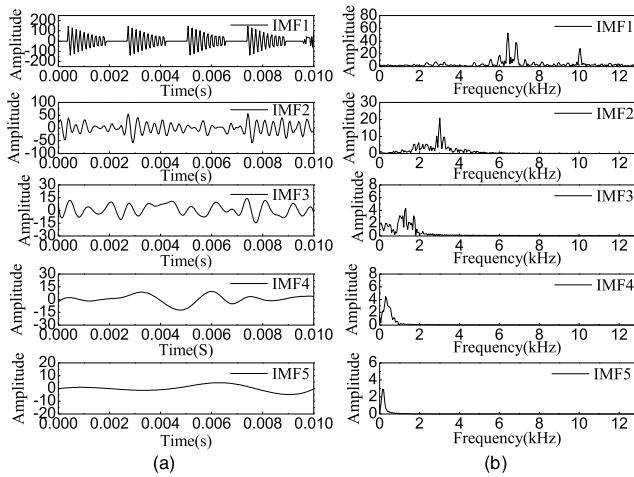


FIGURE 8. Decomposition result of LMD. (a) Time domain. (b) Frequency domain.

TABLE 4. Correlation coefficients between results and original components.

Algorithms	s_1	s_2	s_3
VMD	0.94	0.94	0.96
EWT	0.97	0.78	0.89
EEMD	0.87	0.66	/
LMD	0.88	0.73	/

In order to verify it, the correlation coefficients between results and original components are computed as Table 4.

As shown in Table 4, the correlation coefficients between VMD results and original components are around 0.95. As for EWT, the correlation coefficients for s_2 and s_3 are 0.78 and 0.89 respectively, which are unacceptable values. Although the correlation coefficient between s_1 and IMF2 is 0.97, the result in VMD, i.e. 0.94, has very little influence on recognition. The result also verifies the VMD has better performance in filtering noise. As for EEMD and LMD, they cannot extract high frequency sinusoidal component (i.e. s_3). Besides, although the correlation coefficients of s_1 are high, the frequency domains show that the high frequency sinusoidal component (i.e. s_3) still exists (its energy is low so that has little effect on the computation of correlation coefficients). The correlation coefficients of s_2 are 0.66 and 0.73 respectively, which are much lower than 0.94. In conclusion, the optimized VMD based Fourier spectrum segmentation is a realistic and accurate approach.

C. OPTIMIZED VMD BASED ON POWER SPECTRUM SEGMENTATION

As shown in last section, the Fourier spectrum segmentation can provide accurate and stable parameters for building VMD in order to reduce subjectivity and complexity. However, with in-depth research, author team finds this approach is struggling in dividing approximating frequencies. The f_2 simulated

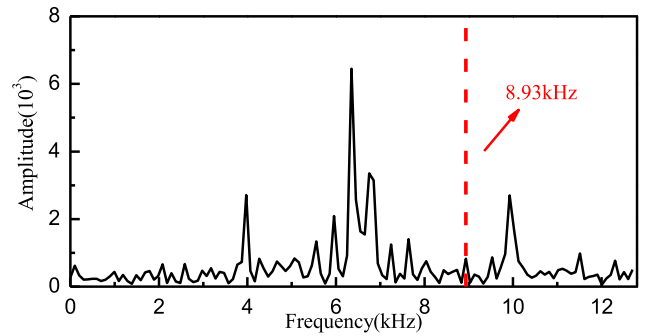


FIGURE 9. Fourier spectrum segmentation of new simulated signal.

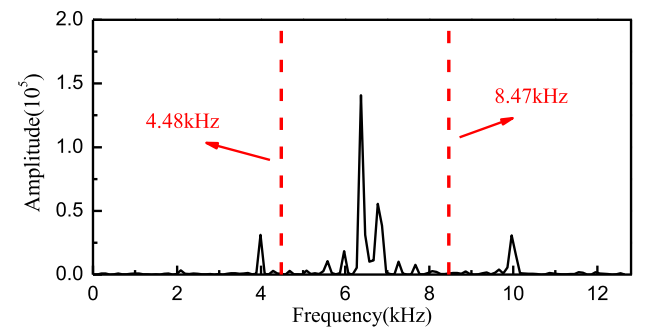


FIGURE 10. Power spectrum segmentation of new simulated signal.

signal shown in equation (16) is shifted from 3000 to 4000 in this section, and the new simulated signal is also divided by Fourier spectrum segmentation. As shown in Fig. 9, only one boundary at 8.93 kHz is obtained, which divides the Fourier spectrum into 2 components.

The reason for that is noise and the approximating frequencies reduce the salience of characteristic components. In order to overcome this problem, a spectrum with similar waveform of Fourier spectrum and salient characteristics should be researched. Based on that, power spectrum is used to divide signal. The power spectrum computed as follows:

Firstly, remove the mean of input signal s :

$$s_{in} = s - s_{mean} \tag{17}$$

where s_{mean} is the mean of s .

Next, compute the Fourier spectrum of s_{in} , and the unilateral Fourier spectrum is $\mathbb{F}(s_{in}) = \{f_1, f_2, \dots, f_{m-1}, f_m\}$. So the power spectrum (PS) is:

$$PS = \left\{ \frac{f_1 * \bar{f}_1}{n}, \frac{f_2 * \bar{f}_2}{n}, \dots, \frac{f_{n-1} * \bar{f}_{n-1}}{n}, \frac{f_n * \bar{f}_n}{n} \right\} \tag{18}$$

where \bar{f}_n is the adjoint of f_n , n is the length of input signal s .

Finally, divide the input signal in scale-space by the power spectrum.

As shown in the description, power spectrum can remain the similar waveform as well as make characteristics stand out. In order to verify this approach, the new simulated signal with $f_2 = 4000$ is divided by it and shown in Fig. 10.

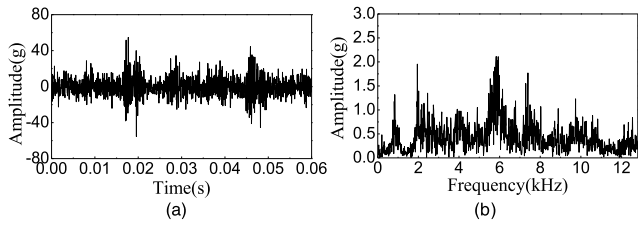


FIGURE 11. Time domain of measured signal. (a) Time domain. (b) Frequency domain.

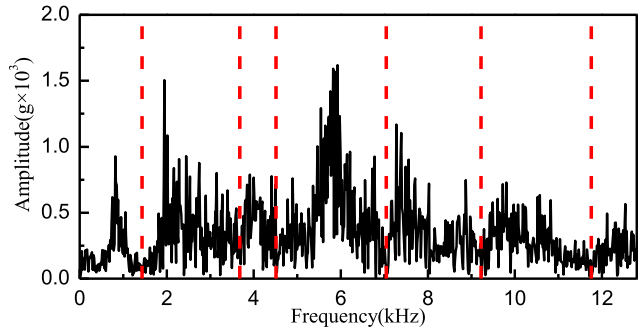


FIGURE 12. Segmentation result of Fourier spectrum.

As shown in Fig. 10, the power spectrum can reduce the influence of noise on segmentation. Two desired boundaries of 4.48 kHz and 8.47 kHz are obtained, which proves this approach is better than Fourier spectrum segmentation.

To verify the applicability of power spectrum segmentation, a measured signal collected from Y-direction (the horizontal which is perpendicular to crankshaft) near the 3rd cylinder at 2100r/min and 100% load is used to test its performance. The duration of the measured signal is an engine cycle and the time domain is shown in Fig. 11.

For this signal, the segmentation result of Fourier spectrum is shown as Fig. 12, 6 boundaries of 1.43 kHz, 3.68 kHz, 4.51 kHz, 7.04 kHz, 9.22 kHz and 11.57 kHz are obtained.

As shown in Fig. 12, 6 boundaries divide the signal into 7 intervals. Among these, the 4th (between 4.51 kHz and 7.04 kHz) and 5th (between 7.04 kHz and 9.22 kHz) intervals is complex. For the 4th interval, it can be identified as a combustion component. According to the simulated signal, periodic impact has wide bandwidth and several extremely close peaks in spectrum, which means this interval is divided appropriately. However, for the 5th interval, there are 2 subsections in it obviously (the segmentation point is near 8 kHz), whose characteristics are different from periodic impact. So the power spectrum is used to divide the signal, and result is shown in Fig. 13.

As shown in Fig. 13, characteristic frequencies stand out significantly, and 7 boundaries of 1.43 kHz, 3.68 kHz, 4.51 kHz, 7.04 kHz, 8.23 kHz, 9.22 kHz and 11.75 kHz are obtained. The interval between 4.51 kHz and 7.04 kHz is still the combustion component. The interval between 7.04 kHz and 9.22 kHz is divided by the boundary of 8.23 kHz into

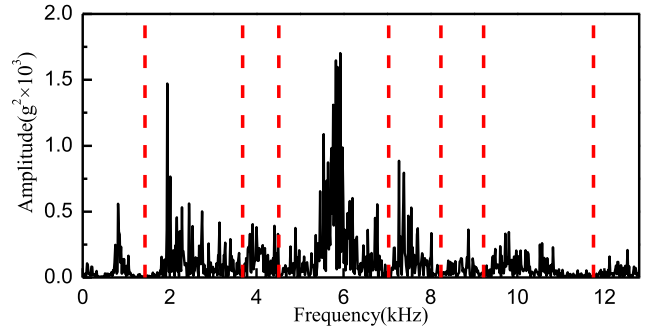


FIGURE 13. Segmentation result of power spectrum.

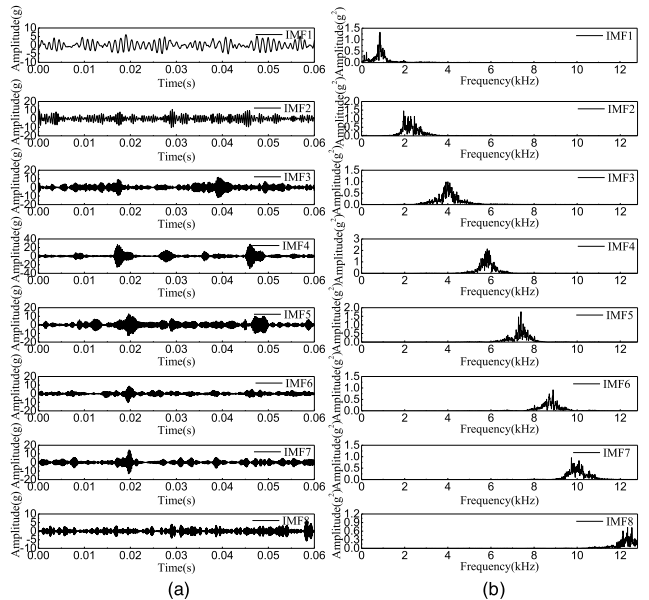


FIGURE 14. Decomposition result of the measured signal. (a) Time domain. (b) Frequency domain.

2 subsections and get more clear segmentation. The result shows the power spectrum has better performance in spectrum analysis and robustness than Fourier spectrum. Based on that, the VMD optimized by power spectrum is used to decompose the measured signal, and the result is shown as Fig. 14.

As shown in Fig. 14, the frequencies of IMFs correspond to the power spectrum segmentation accurately, and ideal narrow bandwidth components with little noise are obtained. Besides, all the IMFs cover the analysis frequency as well as no overlap between them. To explain the advantages of VMD optimized by power spectrum segmentation further, the signal is also decomposed by the VMD optimized by Fourier spectrum segmentation and result is shown in Fig. 15.

As shown in Fig. 15, the component between 8 kHz and 10 kHz is not extracted by the VMD optimized by Fourier spectrum, which means this algorithm has low ability in decomposing components with similar frequencies and may loss important information during decomposition. Overall,

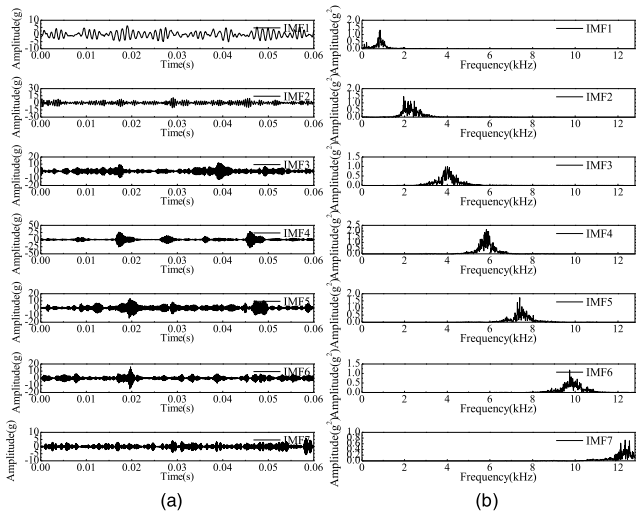


FIGURE 15. Decomposition result of VMD optimized by Fourier spectrum segmentation. (a) Time domain. (b) Frequency domain.

the VMD optimized by power spectrum segmentation is an accurate and advanced approach.

V. EXPERIMENTAL ANALYSIS BASED ON OPTIMIZED EXPECTATION-MAXIMIZATION

A. DATA PREPARATION

The selected vibration signals are collected synchronously from three sensors, two of which are located on the wall of the first and the third cylinders and another one located on the cylinder head cover corresponding to the first cylinder.

The vibration data of each channel is divided into certain length depends on the diesel engine rotating speed and sampling rate. In order to include complete working cycle of diesel engine. The length of segment can be expressed as (19).

$$L = \lceil 60 * fs * 2 / r \rceil \tag{19}$$

where L is the length of segment, fs is sampling frequency, r is rotating speed, the sign “ $\lceil \rceil$ ” represents round up operator.

In order to obtain effective feature subset, original data is decomposed by optimized VMD and the characteristic components are restructured. Nine features are exacted from the restructured signals, the features are shown as follows:

- 1. Minimum value (Min)
- 2. Maximum value (Max)
- 3. Mean value (Mean)
- 4. Mean square value (MS)
- 5. Root mean square (RMS)
- 6. Mean square error (MSE)
- 7. Standard deviation (SD)
- 8. Kurtosis (Kur)
- 9. Margin (Mar)

Considering that signals from 3 channels are analyzed, there are 27 features exacted, where Ch1, Ch2 and Ch3 indicate the first, second and third channels respectively.

The total of 27 features exacted form the signals of 3 channels are shown as follows:

- 1.Min_Ch1
- 2.Min_Ch2
- 3.Min_Ch3
- 4.Max_Ch1
- 5.Max_Ch2
- 6.Max_Ch3
- 7.Mean_Ch1
- 8.Mean_Ch2
- 9.Mean_Ch3
- 10.MS_Ch1
- 11.MS_Ch2
- 12.MS_Ch3
- 13.RMS_Ch1
- 14.RMS_Ch2
- 15.RMS_Ch3
- 16.MSE_Ch1
- 17.MSE_Ch2
- 18.MSE_Ch3
- 19.SD_Ch1
- 20.SD_Ch2
- 21.SD_Ch3
- 22.Kur_Ch1
- 23.Kur_Ch2
- 24.Kur_Ch3
- 25.Mar_Ch1
- 26.Mar_Ch2
- 27.Mar_Ch3

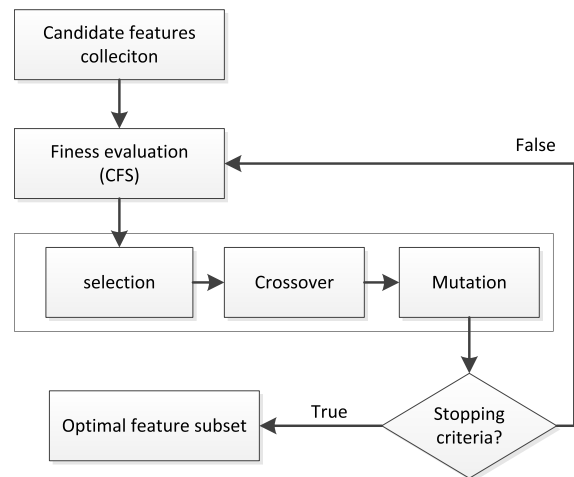


FIGURE 16. GA and fitness function used in feature selection.

B. OPTIMIZED EXPECTATION-MAXIMIZATION ALGORITHM

To improve the adaptability and accuracy of EM, a fitness function of correlation-based feature selection (CFS) approach for GA is introduced to preprocess original data and a cross validation method is used for EM algorithm to select cluster number adaptively.

1) FITNESS FUNCTION OF CORRELATION-BASED FEATURE SELECTION ALGORITHM

The fitness function is a criterion used by attribute subset valuator to measure the worth of an attribute subset. In this article, the CFS evaluator is used as fitness function for GA [33], the process is shown as Fig. 16.

CFS is a filter algorithm that evaluates feature subsets according to correlation based heuristic evaluation function. The bias of the evaluation function is toward feature subsets, which is highly correlated with the class and uncorrelated with each other. The acceptance of a feature will depend on the extent to which it predicts classes in areas of the instance space not already predicted by other features [37]. CFS’s feature subset evaluation function is shown as (20):

$$M_S = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k - 1)\bar{r}_{ff}}} \tag{20}$$

where M_S is the worth of feature subset S containing k features, \bar{r}_{cf} is the mean feature-class correlation ($f \in S$), and \bar{r}_{ff} is the average feature-feature inter correlation. The numerator can be thought as providing an indication to predict the class of feature set, and the denominator indicates the redundancy among features. Equation (20) is the core of CFS and imposes a ranking on feature subsets in the search space of all possible feature subsets.

2) ESTIMATING THE NUMBER OF CLUSTERS USING CROSS-VALIDATION

EM clustering method needs specifying the number of clusters as an input parameter, but in some cases this value is unknown. Many solutions have been proposed to choose the number of clusters automatically, most of them rely on modeling assumptions, but main difficulty in choosing the numbers of clusters is that clustering is fundamentally an “unsupervised” learning problem, meaning that there is no obvious way to use “prediction ability” to drive the model selection [38]. Cross-validation (CV) method is a data-driven approach to estimate the number of clusters, and it is adaptive to the characteristics of data distribution [39]. Specially, in this article, 10-fold cross validation method was adopted, and the details are shown in following steps:

- ①. The number of clusters is set as 1
- ②. The training set is split randomly into 10 folds.
- ③. EM is performed 10 times using the 10 folds the usual CV way.
- ④. The log-likelihood is averaged over all 10 results.
- ⑤. If the log-likelihood increases, the number of clusters is increased by 1 and the program continues at step ②.

The number of folds is fixed to 10, as long as the number of instances in the training set is not smaller 10. If this is the case, the number of folds is set equal to the number of instances [40].

C. EXPERIMENTAL ANALYSIS

Experiments were carried out under the condition of different rail pressures, injection timings and valve clearances, respectively (for detailed parameter settings see TABLE 3). 40 samples at 1600r/min in different working state are selected respectively, and the optimized VMD algorithm was implemented to process the vibration signal acquired synchronously from three vibration sensors at first, then 27 features are extracted from them and preprocessed with feature selection method. Finally, the different working states are classified by EM. The detailed flowchart of data processing is shown as Fig 17.

As shown in Fig 17, the power spectrum and scale-space representation are employed to optimize VMD in order to decompose signals adaptively. Then, the features are extracted from the reconstructed signals. These complex features lead to high computational cost and even reduce the recognition rate, which impels the using of feature selection algorithm. The CFS is selected and GA is used to optimize

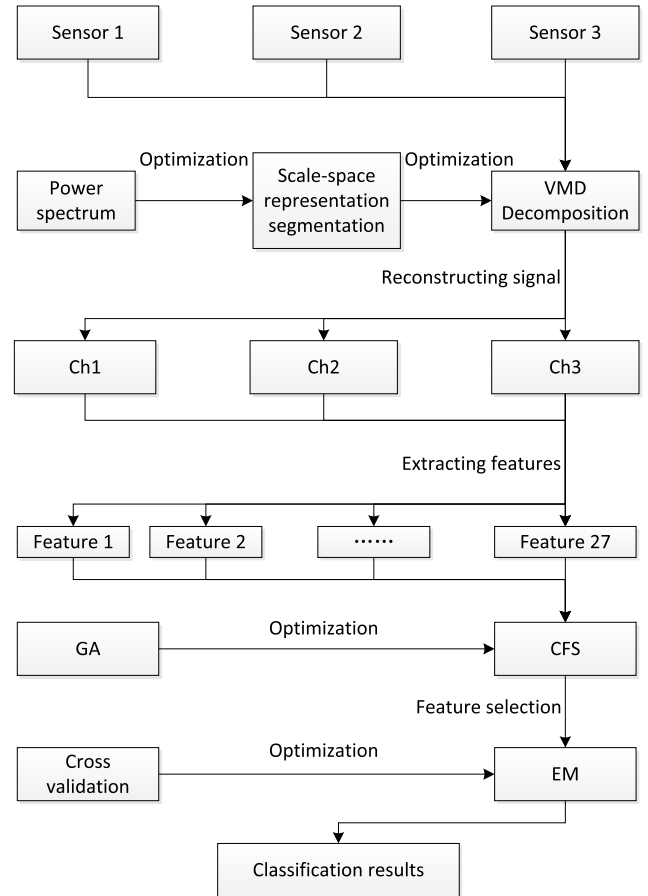


FIGURE 17. Flowchart of data processing.

its performance. After all of these processes, EM optimized by cross validation is used to classify the different working states of engine, in which the cross validation is a method to make the EM selecting the number of clusters adaptively.

To verify the advantages of this scheme, two different unsupervised and adaptive clustering methods, X-means and EM with cross validation, are used to cluster the samples with and without feature selection processing, respectively. For clearly comparison, the classification accuracy is defined as (21):

$$C_{accuracy} = \frac{N_{accuracy}}{N_{original}} \tag{21}$$

where $N_{accuracy}$ is the number of correctly clustered samples; $N_{original}$ is the number of original samples in a certain class. In this experiment, the value of $N_{original}$ is 40.

In this paper, the result of cluster is shown in two-dimensional chart and the classification accuracy is calculated if the number of clusters is correct, otherwise only the result of cluster is shown in two-dimensional chart. In particularly, only MS_Ch1 is shown in charts for limited space because it is a feature exists in every example. The results were shown as follows.

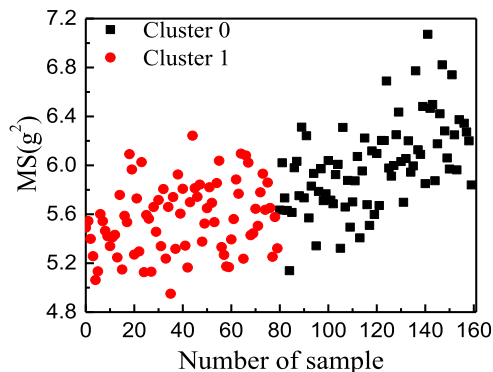


FIGURE 18. X-means without feature selection, number of clusters is 2.

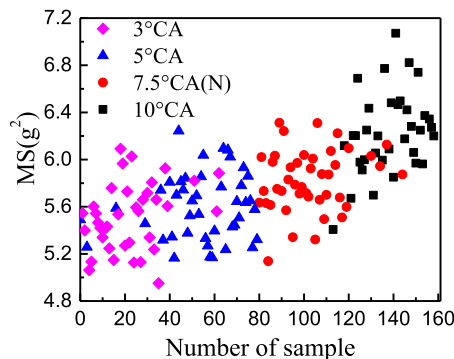


FIGURE 21. EM with feature selection, number of clusters is 4. Classification accuracy: 3° CA: 85% 5° CA: 92.5% 7.5° CA: 95% 10° CA: 85% Average accuracy: 89.38%.

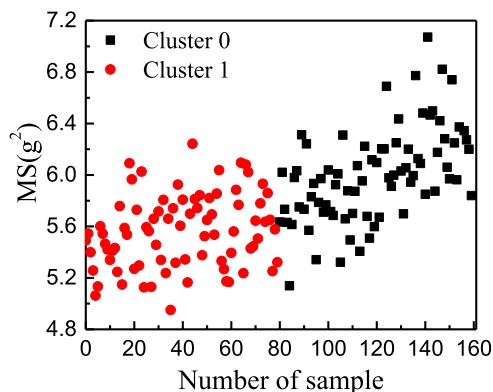


FIGURE 19. X-means with feature selection, number of clusters is 2.

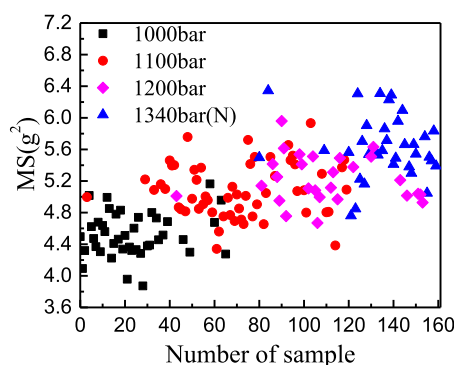


FIGURE 22. X-means without feature selection, number of clusters is 4. Classification accuracy: 1000bar: 87.5% 1100bar: 82.5% 1200bar: 45% 1340bar: 80% Average accuracy: 73.75%.

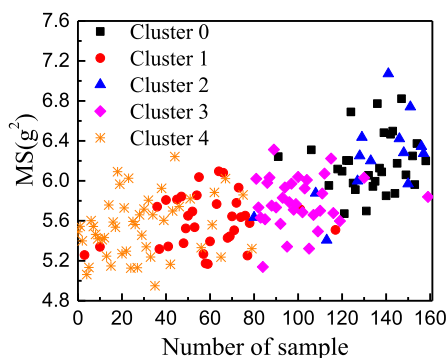


FIGURE 20. EM without feature selection, number of clusters is 5.

1) CLUSTERING RESULT OF DIFFERENT INJECTION TIMING

An optimal subset selected from the original features is shown below, the number of selected features is 15:

- 1.Min_Ch1 3.Min_Ch3 6.Max_Ch3 7.Mean_Ch1
- 8.Mean_Ch2 9.Mean_Ch3 10.MS_Ch1 12.MS_Ch3
- 14.RMS_Ch2 15.RMS_Ch3 16.MSE_Ch1 18.MSE_Ch3
- 21.SD_Ch3 24.Kur_Ch3 27.Mar_Ch3

Four kinds of injection timing parameters are designed in the experiment, so the correct number of clusters is 4.

There are 40 samples for each working state. Experimental clustering result is shown as follows:

As shown in Fig. 18 and Fig. 19, no matter feature selection is carried out or not, X-means cannot get the correct number of clusters. In Fig. 20, EM algorithm also cannot gain the right number of clusters without feature selection. In Fig. 21, after preprocessing of feature selection, the desired cluster result is obtained by EM algorithm.

2) CLUSTERING RESULT OF DIFFERENT RAIL PRESSURES

An optimal subset selected from the original features is shown below, the number of selected features is 13:

- 2.Min_Ch2 3.Min_Ch3 6.Max_Ch 7.Mean_Ch
- 8.Mean_Ch2 9.Mean_Ch 10.MS_Ch 12.MS_Ch3
- 18.MSE_Ch 20.SD_Ch2 21.SD_Ch3 23.Kur_Ch2
- 24.Kur_Ch3

Four kinds of rail pressure parameters are designed in the experiment, and the correct number of clusters is 4. There are 40 samples for each working state. Experimental clustering result is shown as Fig. 22 - Fig. 25.

As shown in Fig. 22 and Fig. 24, without preprocessing of feature selection, both X-means and EM algorithm can get the correct number of clusters and the average classi-

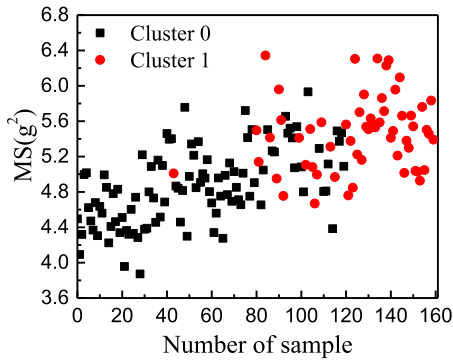


FIGURE 23. X-means with feature selection, number of clusters is 2.

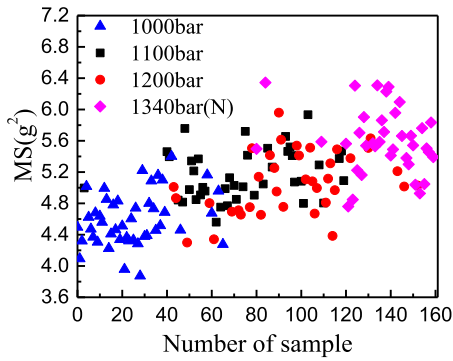


FIGURE 24. EM without feature selection, number of clusters is 4. Classification accuracy: 1000bar: 97.5% 1100bar: 60% 1200bar: 52.5% 1340bar: 85% Average accuracy: 73.75%.

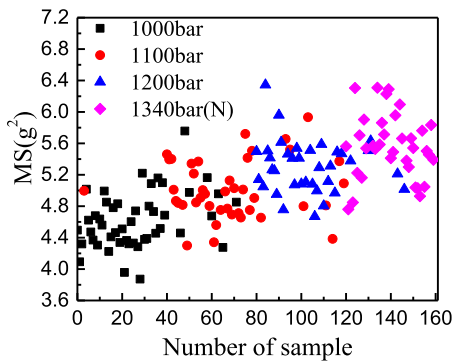


FIGURE 25. EM with feature selection, number of clusters is 4. Classification accuracy: 1000bar: 97.5% 1100bar: 80% 1200bar: 77.5% 1340bar: 87.5% Average accuracy: 85.63%.

fication accuracy of each algorithm is 73.75%. In Fig. 23, X-means cannot gain correct number of clusters with feature-selection preprocessing. In Fig. 25, with feature-selection preprocessing, EM algorithm achieves better result compared with Fig. 22 and Fig. 24, and the average classification accuracy is 85.63%.

3) CLUSTERING RESULT OF DIFFERENT VALVE CLEARANCES

An optimal subset selected from the original features is shown below, the number of selected features is 9:

- 5.Max_Ch2
- 6.Max_Ch3
- 8.Mean_Ch2
- 10.MS_Ch1
- 14.RMS_Ch2
- 16.MSE_Ch1
- 17.MSE_Ch2
- 19.SD_Ch1
- 24.Kur_Ch3

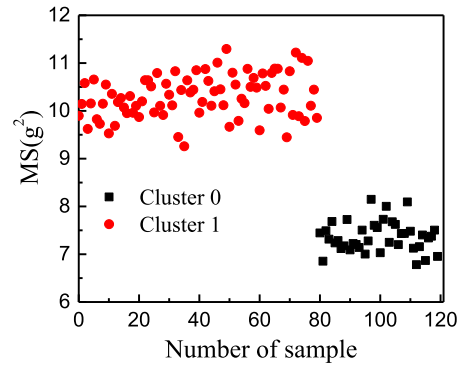


FIGURE 26. X-means without feature selection, number of clusters is 2.

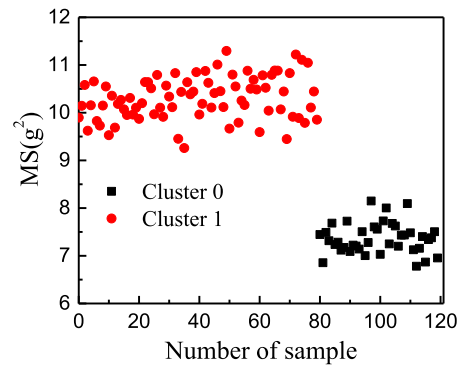


FIGURE 27. X-means with feature selection, number of clusters is 2.

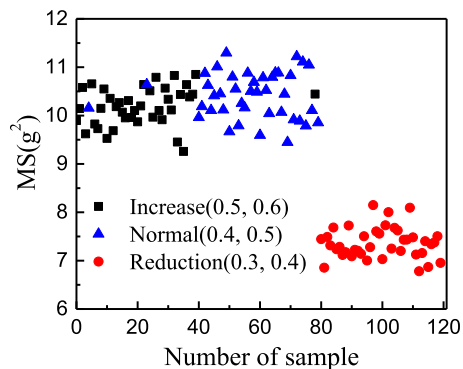


FIGURE 28. EM without feature selection, number of clusters is 4. Classification accuracy: Increase (0.5, 0.6): 95% Normal (0.4, 0.5): 97.5% Reduction (0.3, 0.4): 100% Average accuracy: 97.50%.

The experiment design three group of different valve clearances, and the correct number of clusters is 3. There are 40 samples for each working state. Experimental clustering result is shown as follows:

As shown in Fig. 26 and Fig. 27, no matter feature selection is carried out or not, X-means cannot distinguish three different states of valve clearance. In Fig. 28 and Fig. 29,

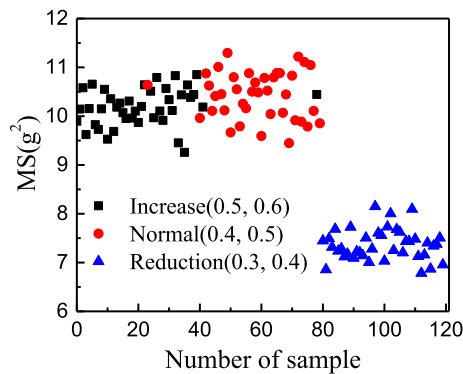


FIGURE 29. EM with feature selection, number of clusters is 4. Classification accuracy: Increase (0.5, 0.6): 97.5% Normal (0.4, 0.5): 95% Reduction (0.3, 0.4): 100% Average accuracy: 97.50%.

three different states are correctly classified by EM algorithm without or with feature selection preprocessing.

It can be deduced from the results that: the EM clustering method with features selection achieves the accurate number of clusters and attains the average classification accuracy of 89.38% for different injection timing parameters. For different rail pressure parameters, X-means and EM both can achieve correct cluster number without features selection, and also achieve the same average classification accuracy of 73.50%. On the contrary, only EM method can get the correct number of clusters after features selection and attains high average classification accuracy of 85.63%. For different valve clearances, X-means method cannot get the right number of clusters no matter with or without features selection. However, EM achieve both correct number of clusters and high average classification accuracy of 97.50%. In summary, the analysis of three types of parameters adjustment experiments in diesel engine validates the EM based on feature selection optimized by genetic algorithm and cross validation has the ability to obtain cluster numbers adaptively and is effective for the classification of different diesel engine working states.

VI. CONCLUSION AND OUTLOOK

In order to recognize states of diesel engine, a novel optimization approach based VMD and EM using multi-channel feature fusion algorithm is proposed:

(1) Segmentation based on scale-space representation is introduced to analyze the frequency components of original vibration signals. Considering low robustness of Fourier spectrum segmentation, power spectrum is used to discriminate frequency components because of its similarity to Fourier spectrum and good performance in noise suppression. Based on that, the decomposition level of VMD can be selected adaptively, which could provide accurate time-frequency analysis results for subsequent recognition.

(2) After extracting features from multi-channel vibration signals, the GA and CFS fitness function are used for features selection to reduce dimensions of data. The cross validation

is introduced to optimized EM algorithm in order to cluster feature subset adaptively.

(3) A diesel engine state recognition scheme based on optimized VMD, feature selection and optimized EM using multi-channel vibration signal is proposed. To prove the advantages of it, classic X-means and the novel scheme are used to analyze different injection timing, rail pressure and valve clearance states. Comparison shows the proposed method has better stability and classification accuracy, and it has certain engineering and theoretical significance.

However, there are some further works need to do:

(1) Quadratic penalty is another important parameter in VMD, which has a big effect on decomposition result, and it is selected as $\alpha = 2200$ by data analysis in this paper. Although obtaining acceptable results, this parameter should be researched further in future work.

(2) There are three main feature selection methods: filter method, wrapper method and embedded method. The feature selection based on GA and CFS fitness function in this paper is one of the filter methods, whose advantage is high efficiency. The shortage is that it is independent of classifier, which results in the selected feature subset maybe not the best. Other feature selection methods will be researched in succeeding work.

REFERENCES

- [1] M. R. Seifi, S. R. Hassan-Beygi, B. Ghobadian, U. Desideri, and M. Antonelli, "Experimental investigation of a diesel engine power, torque and noise emission using water-diesel emulsions," *Fuel*, vol. 166, pp. 392–399, Feb. 2016.
- [2] İ. A. Reşitoğlu, K. Altinişik, and A. Keskin, "The pollutant emissions from diesel-engine vehicles and exhaust aftertreatment systems," *Clean Technol. Environ. Policy*, vol. 17, no. 1, pp. 15–27, Jun. 2014.
- [3] H. M. Nahim, R. Younes, H. Shraim, and M. Ouladsine, "Oriented review to potential simulator for faults modeling in diesel engine," *J. Mar. Sci. Technol.*, vol. 21, no. 3, pp. 533–551, Dec. 2015.
- [4] X. Xu, Z. Liu, J. Wu, J. Xing, and X. Wang, "Misfire fault diagnosis of range extender based on harmonic analysis," *Int. J. Automot. Technol.*, vol. 20, no. 1, pp. 99–108, Feb. 2019.
- [5] Y.-B. Jing, C.-W. Liu, F.-R. Bi, X.-Y. Bi, X. Wang, and K. Shao, "Diesel engine valve clearance fault diagnosis based on features extraction techniques and FastICA-SVM," *Chin. J. Mech. Eng.*, vol. 30, no. 4, pp. 991–1007, May 2017.
- [6] E. Ftoutou and M. Chouchane, "Diesel engine injection faults' detection and classification utilizing unsupervised fuzzy clustering techniques," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 233, no. 16, pp. 5622–5636, May 2019.
- [7] A. Moosavian, M. Khazaei, G. Najafi, M. Khazaei, B. Sakhaei, and S. M. Jafari, "Wavelet denoising using different mother wavelets for fault diagnosis of engine spark plug," *Proc. Inst. Mech. Eng., E, J. Process. Mech. Eng.*, vol. 231, no. 3, pp. 359–370, Jul. 2015.
- [8] F. Bi, T. Ma, and X. Wang, "Development of a novel knock characteristic detection method for gasoline engines based on wavelet-denoising and EMD decomposition," *Mech. Syst. Signal Process.*, vol. 117, pp. 517–536, Feb. 2019.
- [9] Y. Li, P. W. Tse, X. Yang, and J. Yang, "EMD-based fault diagnosis for abnormal clearance between contacting components in a diesel engine," *Mech. Syst. Signal Process.*, vol. 24, no. 1, pp. 193–210, Jan. 2010.
- [10] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
- [11] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method," *Adv. Adapt. Data Anal.*, vol. 2, no. 2, pp. 135–156, Nov. 2011.

- [12] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 60, no. 16, pp. 3999–4010, May 2013.
- [13] J. Gilles and K. Heal, "A parameterless scale-space approach to find meaningful modes in histograms—Application to image and spectrum segmentation," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 12, no. 06, Jan. 2015, Art. no. 1450044.
- [14] D. Wang, K.-L. Tsui, and Y. Qin, "Optimization of segmentation fragments in empirical wavelet transform and its applications to extracting industrial bearing fault features," *Measurement*, vol. 133, pp. 328–340, Feb. 2019.
- [15] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- [16] M. F. Isham, M. S. Leong, M. H. Lim, and Z. A. Ahmad, "Variational mode decomposition: Mode determination method for rotating machinery diagnosis," *J. Vibroeng.*, vol. 20, no. 7, pp. 2604–2621, Nov. 2018.
- [17] B. L. Boada, M. J. L. Boada, and V. Diaz, "Vehicle sideslip angle measurement based on sensor data fusion using an integrated ANFIS and an unscented Kalman filter algorithm," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 832–845, May 2016.
- [18] L. Yi-Bo, WangNing, and ZhouChang, "Based on D-S evidence theory of information fusion improved method," in *Proc. Int. Conf. Comput. Appl. Syst. Modeling (ICCSM)*, Taiyuan, China, Oct. 2010, pp. 416–419.
- [19] Y. Xu, T. Xie, Z. Zhou, Y. Lu, J. Jun, C. Shu, J. Li, Y. Ma, and C. Jiang, "Research of practical expert system for transformer fault diagnosis based on multidimensional data fusion," in *Proc. IEEE PES Asia-Pacific Power Energy Eng. Conf. (APPEEC)*, Oct. 2016, pp. 300–304.
- [20] C. Melendez-Pastor, R. Ruiz-Gonzalez, and J. Gomez-Gil, "A data fusion system of GNSS data and on-vehicle sensors data for improving car positioning precision in urban environments," *Expert Syst. Appl.*, vol. 80, pp. 28–38, Sep. 2017.
- [21] C. D. Stefano, F. Fontanella, and A. S. Freca, "Feature selection in high dimension data by a filter-based genetic algorithm," in *Applications of Evolutionary Computation (Lecture Notes in Computer Science)*, vol. 10199. Cham, Switzerland: Springer, Mar. 2017, pp. 506–521.
- [22] S. Mohammadi, M. Papa, E. Pereyra, and C. Sarica, "Genetic algorithm to select a set of closure relationships in multiphase flow models," *J. Petroleum Sci. Eng.*, vol. 181, Oct. 2019, Art. no. 106224.
- [23] T. Gangavarapu and N. Patil, "A novel filter-wrapper hybrid greedy ensemble approach optimized using the genetic algorithm to reduce the dimensionality of high-dimensional biomedical datasets," *Appl. Soft Comput.*, vol. 81, Aug. 2019, Art. no. 105538.
- [24] R. Malhotra and M. Khanna, "Dynamic selection of fitness function for software change prediction using particle swarm optimization," *Inf. Softw. Technol.*, vol. 112, pp. 51–67, Aug. 2019.
- [25] P. A. Devijver and J. Kittler, "Pattern recognition: A statistical approach," *Image Vis. Comput.*, vol. 3, no. 2, pp. 87–88, May 1985.
- [26] A. K. Das, S. Sengupta, and S. Bhattacharyya, "A group incremental feature selection for classification using rough set theory based genetic algorithm," *Appl. Soft Comput.*, vol. 65, pp. 400–411, Apr. 2018.
- [27] D. Pelleg and A. Moore, "X-menas: Extending K-means with efficient estimation of number of clusters," in *Intelligent Data Engineering and Automated Learning*. Berlin, Germany: Springer, 2000.
- [28] P. Kumar and S. K. Wasan, "Analysis of X-means and global k-means using TUMOR classification," in *Proc. 2nd Int. Conf. Comput. Automat. Eng. (ICCAE)*, Feb. 2010, pp. 832–835.
- [29] H. Jung and B. Seo, "A fast EM algorithm for Gaussian mixtures," *Commun. Korean Stat. Soc.*, vol. 19, no. 1, pp. 157–168, 2012.
- [30] P. Zhao, D. Li, Z. Feng, L. Yao, and G. Jia, "Identification of unknown modes for air-conditioning based on hybrid cluster algorithm," in *Proc. 12th World Congr. Intell. Control Automat. (WCICA)*, Guilin, China, Jun. 2016, pp. 1147–1151.
- [31] J. Chen, X. Zhang, N. Zhang, and K. Guo, "Fault detection for turbine engine disk using adaptive Gaussian mixture model," *Proc. Inst. Mech. Eng., I, J. Syst. Control Eng.*, vol. 231, no. 10, pp. 827–835, Sep. 2017.
- [32] T. Mekaroonkamon and S. Wongsas, "A comparative investigation of the robustness of unsupervised clustering techniques for rotating machine fault diagnosis with poorly-separated data," in *Proc. 8th Int. Conf. Adv. Comput. Intell.*, Chiang Mai, Thailand, Feb. 2016, pp. 165–172.
- [33] F. Tan, X. Fu, Y. Zhang, and A. G. Bourgeois, "A genetic algorithm-based method for feature subset selection," *Soft Comput.*, vol. 12, no. 2, pp. 111–120, May 2007.
- [34] V. Kozeny, "Genetic algorithms for credit scoring: Alternative fitness function performance comparison," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 2998–3004, Apr. 2015.
- [35] Q. Chen, K. Worden, P. Peng, and A. Y. T. Leung, "Genetic algorithm with an improved fitness function for (N)ARX modelling," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 994–1007, Feb. 2007.
- [36] M. Jordan, J. Kleinberg and B. Scholkopf, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006, pp. 423–460.
- [37] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Dept.Comput. Sci., Waikato Univ., Hamilton, New Zealand, 1999.
- [38] D. Ruppert, "The elements of statistical learning: Data mining, inference, and prediction," *J. Amer. Stat. Assoc.*, vol. 99, no. 466, pp. 567–567, Jun. 2004.
- [39] W. Fu and P. O. Perry, "Estimating the number of clusters using cross-validation," *J. Comput. Graph. Statist.*, to be published, doi: 10.1080/10618600.2019.1647846.
- [40] *WEKA-Documentation*, Univ. Waikato, Hamilton, New Zealand, 2018.



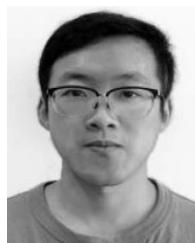
XIAOBO BI received the master's degree in vehicle engineering from the Hebei University of Technology, Tianjin, China. He is currently pursuing the Ph.D. degree in power machine and engineering with the State Key Laboratory of Engines, Tianjin University, Tianjin. He is also a Senior Lecturer with the Hebei University of Technology. His research interests include multiinformation fusion fault diagnosis of ICE and NVH control of vehicle.



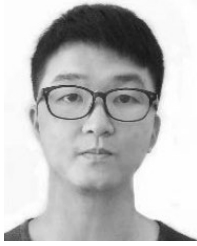
JIANSHENG LIN is currently a Ph.D. Supervisor and a Professor with the Tianjin University of Technology. His current research interests include multibody dynamics simulation and analysis and internal combustion engine working process research.



FENGRONG BI received the Ph.D. degree in power machinery and engineering from Tianjin University, Tianjin, China. He is currently a Ph.D. Supervisor and a Professor with Tianjin University. His current research interests include the control of vibration and noise and the fault diagnosis of vehicles and power machineries, with specific investigation on excitation mechanism, transfer characteristics, assessment, and control methods of NVH.



XIN LI received the master's degree in power machinery and engineering from Tianjin University, Tianjin, China, in 2017, where he is currently pursuing the Ph.D. degree. His main research interests include reliability and weak power fault diagnosis.



DAJIE TANG received the B.S. degree in energy and power engineering and the M.S. degree in power engineering from Tianjin University, Tianjin, China, in 2016 and 2019, respectively, where he is currently pursuing the Ph.D. degree in power machinery and engineering. His current research interests include fault diagnosis of power machinery, vibration, and noise control.



XIAO YANG graduated from Tianjin University, Tianjin, China. He is currently pursuing the Ph.D. degree in power machine and engineering with the State Key Laboratory of Engines, Tianjin University. His current research interests include fault diagnosis and noise reduction of internal combustion engines.



YOUXI WU (Member, IEEE) received the Ph.D. degree in theory and new technology of electrical engineering from the Hebei University of Technology, Tianjin, China. He is currently a Ph.D. Supervisor and a Professor with the Hebei University of Technology. His current research interests include data mining and machine learning. He is a Senior Member of CCF.



PENGFEI SHEN received the B.S. degree in energy and power engineering and the M.S. degree in power engineering from Tianjin University, Tianjin, China, in 2016 and 2019, respectively. His main research interest includes the power machinery's vibration fault diagnosis.

...