Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Application of complete ensemble intrinsic time-scale decomposition and least-square SVM optimized using hybrid DE and PSO to fault diagnosis of diesel engines*

Jun-hong ZHANG[1], Yu LIU[†‡1,2]

([1]*State Key Laboratory of Engines, Tianjin University, Tianjin 300072, China*)

([2]*China Automotive Technology and Research Center, Tianjin 300300, China*)

[†]E-mail: Liuyu2012@tju.edu.cn

Received Oct. 18, 2015; Revision accepted Feb. 17, 2016; Crosschecked Jan. 20, 2017

Abstract: Targeting the mode-mixing problem of intrinsic time-scale decomposition (ITD) and the parameter optimization problem of least-square support vector machine (LSSVM), we propose a novel approach based on complete ensemble intrinsic time-scale decomposition (CEITD) and LSSVM optimized by the hybrid differential evolution and particle swarm optimization (HDEPSO) algorithm for the identification of the fault in a diesel engine. The approach consists mainly of three stages. First, to solve the mode-mixing problem of ITD, a novel CEITD method is proposed. Then the CEITD method is used to decompose the nonstationary vibration signal into a set of stationary proper rotation components (PRCs) and a residual signal. Second, three typical types of time-frequency features, namely singular values, PRCs energy and energy entropy, and AR model parameters, are extracted from the first several PRCs and used as the fault feature vectors. Finally, a HDEPSO algorithm is proposed for the parameter optimization of LSSVM, and the fault diagnosis results can be obtained by inputting the fault feature vectors into the HDEPSO-LSSVM classifier. Simulation and experimental results demonstrate that the proposed fault diagnosis approach can overcome the mode-mixing problem of ITD and accurately identify the fault patterns of diesel engines.

Key words: Diesel; Fault diagnosis; Complete ensemble intrinsic time-scale decomposition (CEITD); Least square support vector machine (LSSVM); Hybrid differential evolution and particle swarm optimization (HDEPSO)

http://dx.doi.org/10.1631/FITEE.1500337 **CLC number:** TK428; TP391

1 Introduction

Recently, vibration analysis has attracted considerable attention and has been widely applied to the fault diagnosis of diesel engines (Tay and Shen, 2003; Li *et al.*, 2010; Wang *et al.*, 2013). The main advantage of using vibration analysis is that the vibration signal can provide a variety of information about the reciprocating and rotating motions, mechanical impacts, and high-speed flow of gas. Moreover, the acquisition of vibration signals is very convenient and nonintrusive (Wang *et al.*, 2008). In this study, we use vibration analysis for the fault diagnosis of diesel engines.

The vibration signals from diesel engines are highly transient and nonstationary (Wang *et al.*, 2008). Therefore, the traditional stationary assumption based Fourier transform (FT) is not suitable for processing such vibration signals. To deal with nonstationary signals, time-frequency analysis methods are often used because they can provide the whole and local characteristics of the signals in time and frequency domains simultaneously. Commonly used time-frequency analysis methods include wavelet transform (WT) (Vong and Wong, 2011), Wigner-Ville

‡ Corresponding author

* Project supported by the National High-Tech R&D Program (863) of China (No. 2014AA041501)

ORCID: Yu LIU, http://orcid.org/0000-0003-0946-4488

distribution (WVD) (Martin and Flandrin, 1985), and empirical mode decomposition (EMD) (Li *et al.*, 2010). Vong and Wong (2011) extracted the ignition fault features of diesel engines by using the WT method, then fed these features into the least-square support vector machine (LSSVM), and obtained a satisfactory fault diagnosis result. Nevertheless, WT is essentially an adjustable-window FT and energy leakage is inevitable. In addition, this method is non-adaptive because it is based on projecting the signals onto a predefined set of base functions (Li *et al.*, 2011). Wang *et al*. (2008) used the WVD method to generate the time-frequency distribution of vibration signals and achieved good performance in diagnosing faults in diesel valve trains. However, the application of WVD is also limited because it suffers from inevitable cross-term interferences when used to process multicomponent signals (Kadambe and Boudreaux-Bartels, 1992). EMD is a self-adaptive time-frequency analysis method by which a multicomponent signal can be decomposed into a set of intrinsic mode functions (IMFs) and a residual signal (Huang *et al.*, 1998; Rilling *et al.*, 2003; Hong *et al.*, 2011). By analyzing each IMF which includes the local characteristic of the signal, the characteristic information of the original signal can be extracted accurately and effectively. However, there still exist some problems to be resolved in EMD, such as the end effects, mode-mixing, and the unexplainable negative frequency (Chen *et al.*, 2011). Recently, Frei and Osorio (2007) proposed a new self-adaptive time-frequency analysis method called intrinsic time-scale decomposition (ITD), by which a multicomponent signal can be decomposed into a set of stationary proper rotation components (PRCs) and a residual signal. Cheng *et al*. (2012) applied ITD to the fault diagnosis of rotating machinery, and their study showed that ITD could effectively extract the nonlinear fault features. Compared with EMD, by using ITD, more information of the signal itself can be used, and the end effects and the unexplainable negative frequency can be relieved. In addition, the computational efficiency is also improved significantly. Nevertheless, when using ITD to process complex signals, the mode-mixing problem, which causes the decomposition results to lose their physical meaning and provides false characteristic information, is still a major drawback (Lin, 2012). To alleviate the mode-mixing problem of EMD, Wu and Huang (2009) proposed an improved version of EMD, called ensemble empirical mode decomposition (EEMD). EEMD is based on the noise-assisted analysis technique, by which the mode-mixing problem can be eliminated automatically. Zhang *et al.* (2015) applied EEMD to the fault diagnosis of rolling bearing, and compared EEMD with the original EMD method. Their results indicated that EEMD is effective in solving the mode-mixing problem. Similar conclusions were obtained by applying EEMD to the fault diagnosis of the fuel system of a diesel engine and rotor system (Lei *et al.*, 2009; Wang *et al.*, 2013). However, even though EEMD was shown to be effective in many applications, it also created two major problems. One is that the residual noise cannot be completely canceled out; the other is that different realizations of signal plus noise may produce different numbers of modes, which will make the final averaging very difficult to implement (Torres *et al.*, 2011; Xue *et al.*, 2015). Therefore, to alleviate the mode-mixing problem of ITD, simply using the framework of EEMD cannot produce a satisfactory result. To address this problem, a novel complete ensemble intrinsic time-scale decomposition (CEITD) method is proposed in this paper for processing the vibration signals of diesel engines.

Apart from signal processing, pattern recognition is another important part of fault diagnosis. Conventional statistical pattern recognition methods, such as decision tree (Chen *et al.*, 2004) and rough set theory (Tay and Shen, 2003; Huang *et al.*, 2015), once provided satisfactory results in fault diagnosis. However, each pattern recognition method has its own insurmountable problems. For decision tree, professional knowledge is required for the users, and its computation load is relatively large. Therefore, it is not suitable for solving complex problems. The application of rough set theory is also limited because it is not stable and is very susceptible to noise interference. Artificial neural network (ANN) has strong nonlinear fitting ability and high reliability, and it has been widely used in fault diagnosis in the past decades. Moosavian *et al*. (2013) applied ANN to the fault diagnosis of a main engine journal-bearing and compared it with the K-nearest neighbor (K-NN) method. The results showed that the performance of ANN was better than that of K-NN.

Martínez-Martínez *et al*. (2015) predicted the status of the rotary components by using ANN, and used the genetic algorithm (GA) to fit the ANN weights and biases. The results indicated that the proposed method could achieve a mean success rate of 92.96%. Although it has many successful applications, ANN is still not an ideal pattern recognition method. This is because, when the number of samples is insufficient, the inherent deficiencies, such as local optimal solution and overfitting, become significant issues (Huang *et al.*, 2011). All these problems of ANN can be relieved by support vector machine (SVM) (Vapnik, 1999). SVM is based on the structural risk minimization principle, which can minimize an upper bound on the expected risk, so that it can solve the problems of overfitting and local optimal solution. Moreover, SVM implements classification by using a separating hyperplane that is determined by a few samples called support vectors, so that it is applicable to small sample cases. However, the computational complexity of training SVM is rather high, and the solving difficulty of the quadratic programming problem increases with the increase of the number of training samples (Suykens and Vandewalle, 1999). To make SVM more practical in engineering applications, Suykens and Vandewalle (1999) proposed a modified version of SVM called LSSVM. In contrast to SVM, the training process of LSSVM is highly simplified, since a linear problem is resolved instead of a quadratic programming problem in the SVM case. Jiang *et al*. (2015) proposed a novel gearbox fault diagnosis method, which uses the scaling crossover characteristics of signals for self-adaptive feature extraction and LSSVM for pattern recognition. Their method was sensitive to multiple-fault vibration data of a gearbox with similar fault patterns and had a better performance than other methods. In addition, LSSVM has been successfully employed in the fault diagnosis of other machineries because of its good generalization ability and less computational cost (Vong and Wong, 2011; Li *et al.*, 2015; Su *et al.*, 2015). From these applications, we can find that the performance of LSSVM depends heavily on the selection of the regularization parameter and kernel parameter. In recent decades, research on how to determine the best parameters of LSSVM has been a hot topic (Cheng *et al.*, 2013; Xu and Chen, 2013). Commonly used parameter optimization algorithms include GA (Xie *et al.*, 2009), particle swarm algorithm (PSO) (Xu and Chen, 2013), and differential evolution (DE) (Cheng *et al.*, 2013). GA was the most widely used parameter optimization algorithm in the past decades. However, its computation cost is relatively high for adopting the bit-string encoding strategy. In addition, GA often falls into the local optimal solution. The PSO algorithm simulates the social behavior of bird flocks foraging, and it is simple to implement and easy to compute. However, it also suffers from the problem of being trapped into a local optimal solution (Ardia *et al.*, 2011). As an improved version of GA, DE has attracted much attention recently for using floating-point instead of bit-string encoding of population members and arithmetic operations instead of logical operations in mutation (in contrast to classic GA) (Ardia *et al.*, 2011). However, the performance of DE is sensitive to the selection of the mutation strategy (Mallipeddi *et al.*, 2011). Usually, a reliability mutation strategy is selected to guarantee the global searching ability, whereas this strategy often leads to slow convergence. In other words, different parameter selection methods have their own advantages and disadvantages. To realize advantageous complementarities between DE and PSO algorithms, a hybrid differential evolution and particle swarm optimization (HDEPSO) algorithm is proposed in this paper for the parameter optimization of LSSVM.

2 Complete ensemble intrinsic time-scale decomposition

2.1 Intrinsic time-scale decomposition method

For details of ITD, we refer readers to Frei and Osorio (2007). In the original ITD method, the baseline is obtained by using linear interpolation, but linear interpolation is not accurate enough for processing the vibration signals of diesel engines. Therefore, in this study, Hermite interpolation is used to fit the baseline of ITD.

2.2 Mode-mixing in ITD

The appearance of mode-mixing is a major shortcoming of the ITD method. Mode-mixing is defined as a random PRC consisting of fluctuation of different timescales or oscillations with similar scales residing in different PRCs. The mode-mixing

problem, which usually results from signal intermittency, would not only make the corresponding PRC lose its physical meaning but also cause serious aliasing in the time-frequency distribution (Lei *et al.*, 2009; Lin, 2012). To illustrate the mode-mixing problem in ITD, a simulated signal is considered in this section. The simulated signal $x(t)$ has a low-frequency sinusoidal wave $x_1(t)$ with unit amplitude as the fundamental part. At two middle crests of the low-frequency wave, an intermittent high-frequency oscillation $x_2(t)$ with an amplitude of 0.2 rides on the fundamental. The waveforms of the simulated signal $x(t)$ and its two components $x_1(t)$ and $x_2(t)$ are shown in Fig. 1. The ITD method is used to decompose signal $x(t)$ into several PRCs and a residual signal.

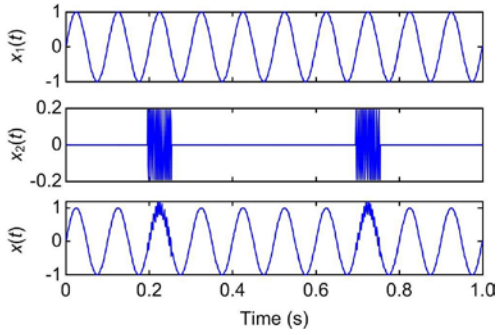

Fig. 1 Simulated signal $x(t)$ and its two components $x_1(t)$ and $x_2(t)$

The decomposition results are shown in Fig. 2, which shows that mode-mixing occurs between PRC_1 and PRC_2. Both the sinusoidal wave $x_1(t)$ and the high-frequency intermittent oscillation $x_2(t)$ are decomposed into PRC_1. Moreover, part of the sine wave $x_1(t)$ is decomposed into PRC_2. The mode-mixing problem of ITD makes the PRCs lose their physical meaning, and these PRCs cannot represent the characteristics of their corresponding true components accurately.

2.3 Ensemble intrinsic time-scale decomposition based on noise-assisted analysis

To solve the mode-mixing problem of EMD, many researchers employed the noise-assisted analysis technique (Lei *et al.*, 2009; Wu and Huang, 2009; Torres *et al.*, 2011; Xue *et al.*, 2015; Zhang *et al.*, 2015). The motivation of this method is that the white noise could provide a uniformly distributed scale in the time-frequency space. When a signal is added to

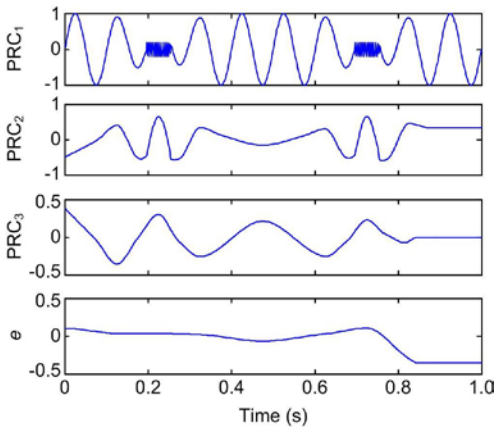

Fig. 2 Decomposition results PRC_1, PRC_2, and PRC_3 and the residue e of the simulated signal $x(t)$ shown in Fig. 1 by using ITD

this white-noise background, the components in different scales of the signal can automatically associate with the similar scales of the background. Therefore, the mode-mixing problem can be controlled by the reference established by the white noise (Wu and Huang, 2009). Inspired by using the noise-assisted analysis technique for EMD, the ITD method is used to replace EMD in the EEMD method, and this method is named ensemble intrinsic time-scale decomposition (EITD). The detailed EITD method is given as follows:

Step 1: Initialize the number of ensembles M, the amplitude A of the added white noise, and i=1.

Step 2: Add a random white noise with amplitude A to the targeted signal $x(t)$, and a new mixing signal $y_i(t)$ can be obtained:

$$y_i(t) = x(t) + n_i(t), \tag{1}$$

where $y_i(t)$ is the noise-added signal of the ith realization and $n_i(t)$ is the ith added white noise.

Step 3: Decompose signal $y_i(t)$ into p PRCs by using the ITD method:

$$y_i(t) = \sum_{j=1}^{p} \mathrm{PRC}_{ij}(t) + u_i(t), \tag{2}$$

where $\mathrm{PRC}_{ij}(t)$ denotes the jth PRC of the ith realization.

Step 4: If $i<M$, repeat steps 2–4 with $i=i+1$; else, go to step 5.

Step 5: Regard the ensemble $\overline{\mathrm{PRC}_j(t)}$ of the M

realizations for each $\mathrm{PRC}_j(t)$ as the final decomposition result:

$$\overline{\mathrm{PRC}_j(t)}=\sum_{i=1}^{M}\frac{\mathrm{PRC}_{ij}(t)}{M},\ i=1,2,\ldots,M,\ j=1,2,\ldots,K, \tag{3}$$

where K is the number of PRCs.

To demonstrate EITD's performance in overcoming the mode-mixing problem, the simulated signal $x(t)$ in Section 2.2 is decomposed again by using EITD with the ensemble number 100 and the added noise amplitude 0.05 times the standard deviation of the simulated signal $x(t)$. The decomposition results are shown in Fig. 3.

Comparing Figs. 1 and 3, it can be observed that the intermittent high-frequency oscillation $x_2(t)$ is decomposed into PRC_1 perfectly, while the sinusoidal wave $x_1(t)$ is decomposed into PRC_3, PRC_4, and PRC_5. Compared with the ITD method, EITD has alleviated the mode-mixing problem to some extent. However, further improvement is still needed to completely eliminate the problem.

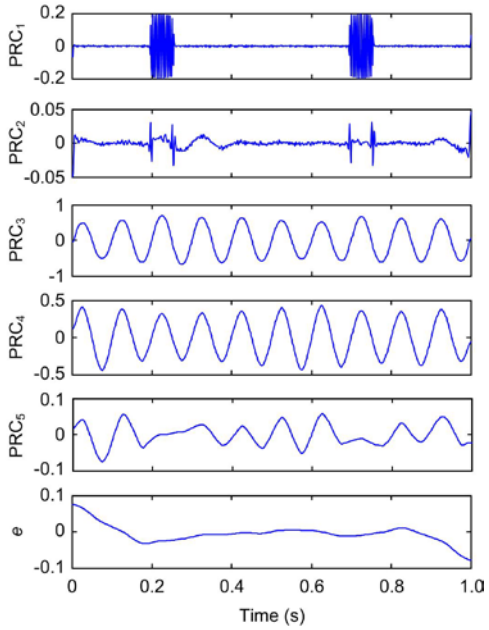

Fig. 3 Decomposition results PRC_1, PRC_2, …, PRC_5 and the residue e of the simulated signal $x(t)$ shown in Fig. 1 by using EITD

2.4 Complete ensemble intrinsic time-scale decomposition

Besides mode-mixing alleviation, it can be found that the mixing signals of each realization in EITD are different, and they are decomposed independently. In other words, there is no connection between different realizations during the EITD. This situation may cause the number of PRCs to be not equal in different realizations and make the final averaging very difficult to implement. In addition, since white noise has a strong randomness, even if a large number of ensembles is selected, the noise still cannot be completely canceled out in the ensemble-averaging process (Xue *et al.*, 2015). Taking into account these drawbacks, a CEITD method is proposed. Define the operator $H_j(\cdot)$ which, given a signal, produces the jth component obtained by ITD. The detailed CEITD method is described as follows:

Step 1: Initialize different realizations of random white noise $n_i(t)$ (i=1, 2, ..., M) with amplitude A, where M is an even number and $n_i(t)=-n_{M/2+i}(t)$ (i=1, 2, ..., M/2). In this way, the white noise is added in pairs (one positive and one negative) to the targeted signal $x(t)$.

Step 2: For i=1, 2, ..., M, decompose each $x(t)+n_i(t)$ by using ITD until the first PRC is obtained. Calculate the first PRC of CEITD:

$$\overline{\mathrm{PRC}_1(t)}=\sum_{i=1}^{M}\frac{\mathrm{PRC}_1^i(t)}{M}, \tag{4}$$

where $\mathrm{PRC}_1^i(t)$ is the first PRC of the ith mixing signal $x(t)+n_i(t)$.

Step 3: Calculate the first residue:

$$r_1(t)=x(t)-\overline{\mathrm{PRC}_1^i(t)}. \tag{5}$$

Step 4: Obtain the first PRC of $r_1(t)+\alpha_1 H_1(n_i(t))$ by using ITD and define the second PRC of CEITD as

$$\overline{\mathrm{PRC}_2(t)}=\sum_{i=1}^{M}\frac{H_1(r_1(t)+\alpha_1\cdot H_1(n_i(t)))}{M}. \tag{6}$$

Step 5: For k=2, 3, …, K, calculate the kth residue:

$$r_k(t)=r_{k-1}(t)-\overline{\mathrm{PRC}_k^i(t)}. \tag{7}$$

Step 6: Obtain the first PRC of $r_k(t)+\alpha_k\cdot H_k(n_i(t))$ by ITD and define the (k+1)th PRC of CEITD as

$$\overline{\mathrm{PRC}_{k+1}(t)}=\sum_{i=1}^{M}\frac{H_1(r_k(t)+\alpha_k\cdot H_k(n_i(t)))}{M}, \tag{8}$$

where the coefficients $\alpha_k=\varepsilon\cdot\mathrm{std}(r_k(t))/\mathrm{std}(H_k(n_i(t)))$ are used to ensure a constant signal-to-noise ratio (SNR) between the added noise and the residue to which the noise is added, and ε is the ratio between noise amplitude and the standard deviation of signal $x(t)$.

Step 7: Go to step 5 with $k=k+1$, and repeat steps 5–7 until $r_K(t)$ meets the termination condition of ITD. Finally, signal $x(t)$ can be expressed as

$$x(t)=\sum_{k=1}^{K}\overline{\mathrm{PRC}_k(t)}+r_K(t). \tag{9}$$

To demonstrate CEITD's performance in overcoming the mode-mixing problem, the simulated signal $x(t)$ in Section 2.2 is decomposed again by using CEITD with the ensemble number 100 and the added noise amplitude 0.05 times the standard deviation of signal $x(t)$. The decomposition results are shown in Fig. 4.

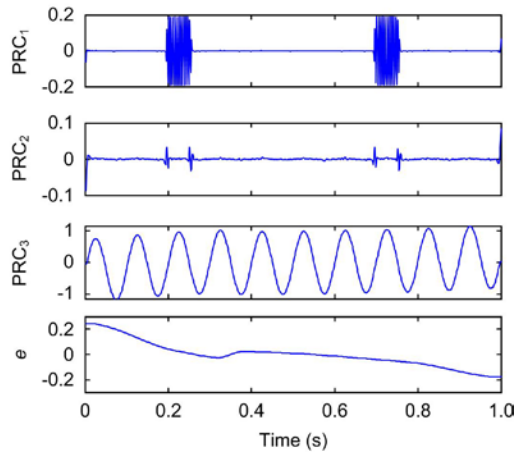

Fig. 4 Decomposition results PRC_1, PRC_2, and PRC_3 and the residue e of the simulated signal $x(t)$ shown in Fig. 1 by using CEITD

Comparing Figs. 1 and 4, it can be found that the two components contained in signal $x(t)$ are decomposed into two PRCs perfectly. PRC_1 and PRC_3 are in accordance with $x_2(t)$ and $x_1(t)$, respectively. Therefore, it can be concluded that the CEITD method can solve the mode-mixing problem of ITD and that it is better than the EITD method. In addition, PRC_2 is a false component caused by the sudden change of the extreme point in the signal, and it can be eliminated by calculating the correlation coefficients between the PRCs and the simulated signal $x(t)$.

3 Least-square SVM with hybrid differential evolution and particle swarm optimization

3.1 Least-square SVM

Given a training set $\{\boldsymbol{x}_i,y_i\}_{i=1}^N$, where $\boldsymbol{x}_i\in\mathbb{R}^n$ is the ith training sample and $y_i\in(0, 1)$ is the class label of $\boldsymbol{x}_i$. To classify the training set, the LSSVM (Suykens and Vandewalle, 1999) approach constructs the optimal separating hyperplane in the form of

$$y(\boldsymbol{x})=\mathrm{sign}(\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x})+b), \tag{10}$$

where $\varphi(\cdot)$ is a nonlinear function that maps the input data into a higher dimensional feature space, $\boldsymbol{w}$ is the weight vector of the separating hyperplane, and b is the basis term.

Parameters $\boldsymbol{w}$ and b can be obtained by solving the optimization problem as follows:

$$\begin{aligned}&\min_{\boldsymbol{w},e}J(\boldsymbol{w},\xi)=\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+\frac{\gamma}{2}\sum_{i=1}^{N}\xi_i^2\\&\text{s.t.}\quad y_i[\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+b]=1-\xi_i,\ i=1,2,\ldots,N,\end{aligned} \tag{11}$$

where γ is the regularization parameter and ξ_i is the nonnegative error for misclassification. For the minimization function in Eq. (11), the first term in the right-hand side stands for the minimization of the Vapnik-Chervonenkis (VC) dimension, while the second term in the right-hand side minimizes the training errors. Define the Lagrangian function as follows:

$$\begin{aligned}L(\boldsymbol{w},b,\xi,\alpha)&=J(\boldsymbol{w},\xi)\\&-\sum_{i=1}^{N}\mu_i\{y_i[\boldsymbol{w}^{\mathrm{T}}\varphi(\boldsymbol{x}_i)+b]-1+\xi_i\},\end{aligned} \tag{12}$$

where μ_i is the Lagrange multiplier and only the support vectors obtain nonzero μ_i. The optimal solution of Eq. (12) must satisfy $\partial L/\partial\boldsymbol{w}=0$, $\partial L/\partial b=0$, $\partial L/\partial\xi_i=0$, and $\partial L/\partial\mu_i=0$. After the elimination of $\boldsymbol{w}$ and $\boldsymbol{\xi}$, the following linear system is obtained:

$$\begin{bmatrix}0 & \boldsymbol{Y}^{\mathrm{T}}\\ \boldsymbol{Y} & \boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}+\gamma^{-1}\boldsymbol{I}\end{bmatrix}\begin{bmatrix}b\\ \boldsymbol{\mu}\end{bmatrix}=\begin{bmatrix}0\\ \boldsymbol{1}\end{bmatrix}, \tag{13}$$

where

$$\boldsymbol{Z}=\begin{bmatrix}\varphi(\boldsymbol{x}_1)^{\mathrm{T}} & y_1\\ \varphi(\boldsymbol{x}_2)^{\mathrm{T}} & y_2\\ \vdots & \vdots\\ \varphi(\boldsymbol{x}_N)^{\mathrm{T}} & y_N\end{bmatrix},$$

$\boldsymbol{Y}=[y_1, y_2, \ldots, y_N]^{\mathrm{T}}$, $\boldsymbol{\mu}=[\mu_1, \mu_2, \ldots, \mu_N]^{\mathrm{T}}$, and $\mathbf{1}=[1, 1, \ldots, 1]^{\mathrm{T}}$.

Then, a kernel function satisfying the Mercer condition is introduced:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j)=\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\varphi(\boldsymbol{x}_j). \tag{14}$$

Finally, the classification function of LSSVM is given by

$$y(\boldsymbol{x})=\operatorname{sgn}\left(\sum_{i=1}^{N}\mu_i y_i K(\boldsymbol{x}, \boldsymbol{x}_k)+b\right). \tag{15}$$

Commonly used kernel functions include the polynomial function, radial basis function (RBF), and sigmoid kernel function, among which the RBF kernel function has been proven to be very effective (Vong and Wong, 2011; Xu and Chen, 2013; Su *et al.*, 2015). Therefore, in this study, the RBF kernel function is adopted as the kernel function of LSSVM.

3.2 Parameter optimization of LSSVM based on HDEPSO

The performance of LSSVM depends heavily on the selection of two parameters. One is the regularization parameter γ, by which the trade-off between the training error minimization and smoothness of the classification function can be determined. The other is the kernel parameter σ, which has an important influence on the distribution of the data in a high-dimensional space. To obtain a satisfactory classification model of LSSVM, in this section, a novel HDEPSO algorithm is proposed to optimize the regularization parameter and kernel parameter of LSSVM.

3.2.1 Particle swarm optimization

For an optimization problem in a D-dimensional space, the PSO algorithm is initialized with a group of random particles, and each particle is a possible solution. To find the optimal solution, each particle adjusts its searching velocity and direction according to the experience of the particle population and its own experience, and this process will be repeated until a terminal condition is met. The velocity and position updating formulas are as follows (Eberhart and Kennedy, 1995):

$$\boldsymbol{v}_i^{m+1}=w\cdot\boldsymbol{v}_i^m+c_1 r_1(\boldsymbol{p}_i^m-\boldsymbol{x}_i^m)+c_2 r_2(\boldsymbol{p}_g^m-\boldsymbol{x}_i^m), \tag{16}$$

$$\boldsymbol{x}_i^{m+1}=\boldsymbol{x}_i^m+\boldsymbol{v}_i^{m+1}, \tag{17}$$

where $\boldsymbol{x}_i^m=(x_{i1}^m, x_{i2}^m, \ldots, x_{iD}^m)$ is the position of the ith particle at the mth iteration, $\boldsymbol{v}_i^m=(v_{i1}^m, v_{i2}^m, \ldots, v_{iD}^m)$ is the velocity of the ith particle at the mth iteration, $\boldsymbol{p}_i^m=(p_{i1}^m, p_{i2}^m, \ldots, p_{iD}^m)$ is the optimal position that the ith particle has reached, $\boldsymbol{p}_g^m=(p_{g1}^m, p_{g2}^m, \ldots, p_{gD}^m)$ is the optimal position that the population has reached, w is the inertia weight factor, c_1 is the cognition learning factor, c_2 is the social learning factor, and r_1 and r_2 are two random numbers in the range [0, 1]. Previous studies have shown that the inertia weight factor w, which is used to control the impact of previous velocity on current iteration, is a very important parameter for the PSO algorithm (Xu and Chen, 2013). A large w can improve the global searching ability but with worse local searching ability and lower convergence speed, while a relatively small w can improve the local searching ability and convergence speed but with more likelihood of falling into a local optimum. Here, a variable inertia weight is used to ensure strong global searching ability in the prophase of the whole searching process and strong local searching ability and high convergence speed in the anaphase of the whole searching process:

$$w_m=w_{\max}-(w_{\max}-w_{\min})\cdot\left(\frac{m-1}{t}\right)^2, \tag{18}$$

where m is the index of the current generation, t is the maximum number of generations, $w_{\max}$ is the maximum inertia weight factor, and $w_{\min}$ is the minimum inertia weight factor.

3.2.2 Differential evolution

For an optimization problem in a D-dimensional space, the DE algorithm randomly initializes a population of N_P individuals within the search space (Storn and Price, 1997). The ith individual of the population $\boldsymbol{x}_i$, the so-called target, can be expressed as $\boldsymbol{x}_i=(x_{i,1}, x_{i,2}, \ldots, x_{i,D})$, $i=1, 2, \ldots, N_P$.

After initialization, DE employs the mutation operation to produce a mutant vector $\boldsymbol{v}_i$ for each individual in the population. There are many strategies for the mutation operation, among which the following strategy is frequently used:

$$\boldsymbol{v}_i = \boldsymbol{x}_{i_1} + F \cdot (\boldsymbol{x}_{i_2} - \boldsymbol{x}_{i_3}), \tag{19}$$

where indices i_1, i_2, and i_3 are mutually exclusive integers randomly generated within range $[1, N_P]$. For each mutant vector, these indices are randomly generated once. F is a positive scaling factor, used to control the influence of the difference vector.

After the mutation phase, the binomial crossover operation is applied to each pair of the mutant vector $\boldsymbol{v}_i$ and its corresponding target vector $\boldsymbol{x}_i$ to generate a trial vector $\boldsymbol{u}_i=(u_{i,1}, u_{i,2}, \ldots, u_{i,D})$:

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{rand}_b \le \text{CR} \ \text{ or } \ j = \text{rand}_r, \\ x_{i,j}, & \text{otherwise}, \end{cases} \quad j = 1, 2, \ldots, D, \tag{20}$$

where rand_b is a random number in the range [0, 1], CR is the crossover probability, which controls the fraction of parameter values that are copied from the mutant vector, and rand_r is a randomly chosen integer in the range $[1, D]$, used to ensure that the trial vector $\boldsymbol{u}_i$ has some difference with the target vector $\boldsymbol{x}_i$.

If an element of the trial vector is found to exceed the bounds after mutation and crossover, it will be reset in the pre-specified range. Then, the objective function values for all trial vectors are evaluated. If a trial vector $f(\boldsymbol{u}_i)$ has an objective function value which is not larger than that of the corresponding target vector $f(\boldsymbol{x}_i)$, it will replace the corresponding target vector in the population for the next generation; otherwise, the target vector $\boldsymbol{x}_i$ will be retained. The selection operation is expressed as follows:

$$\boldsymbol{x}_i = \begin{cases} \boldsymbol{u}_i, & f(\boldsymbol{u}_i) \le f(\boldsymbol{x}_i), \\ \boldsymbol{x}_i, & \text{otherwise}. \end{cases} \tag{21}$$

The mutation, crossover, and selection operations will be repeated until a termination condition is met.

3.2.3 Hybrid differential evolution and particle swarm optimization

Targeting the commonality of iterative optimization of DE and PSO, and to realize advantageous complementarities, a hybrid optimization algorithm called HDEPSO is proposed. The detailed HDEPSO algorithm is as follows (Fig. 5):

Step 1: Initialize the control parameters of DE and PSO, including the maximum number of generations, scaling factor of mutation operation, crossover probability, cognition learning factor, social learning factor, maximum inertia weight factor, and minimum inertia weight factor.

Step 2: Initialize the population of DE, and use this population also for PSO. Initialize the velocities of the particles in the PSO algorithm.

Step 3: Calculate the individual fitness and select the optimal individual of DE. Calculate the individual fitness and find out the optimal position of each particle and the optimal position of the population of PSO.

Step 4: Compare the optimal solution of DE and PSO in each generation. If the optimal solution of PSO has an objective function value which is not larger than that of DE, the optimal solution of DE will be regarded as the overall optimal solution, and the worst particle of PSO will be replaced by the best individual of DE; otherwise, the optimal solution of PSO will be regarded as the overall optimal solution, and the worst individual of DE will be replaced by the best particle of PSO.

Step 5: Perform the mutation, crossover, and selection operations on the population of DE. Update the velocity and position of each particle in the PSO algorithm.

Step 6: Repeat steps 3–5 generation after generation until a termination criterion (reaching the maximum number of generations) is satisfied. Then the overall optimal solution is outputted as the optimization results of the HDEPSO algorithm.

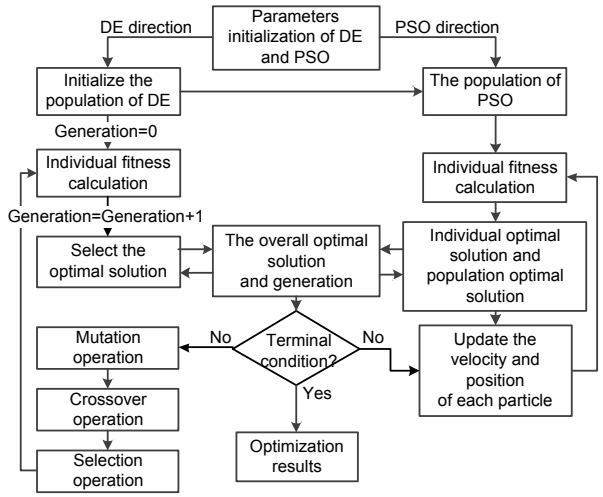

Fig. 5 Flowchart of the hybrid differential evolution and particle swarm optimization (HDEPSO) algorithm

4 Feature extraction and the fault diagnosis approach based on CEITD and LSSVM

4.1 Feature extraction

4.1.1 Singular value decomposition

On the basis of matrix theory, singular values can reflect the nature and characteristics of matrix. However, its application is often limited in fault diagnosis because the reconstruction parameters, such as embedding dimension and lag time, must be determined before the singular value decomposition (SVD) technique is used. Fortunately, after the decomposition of CEITD, the initial feature vector matrices of the SVD technique are formed automatically by the PRCs. Therefore, the characteristics of the fault signal can be described by the singular values of the PRCs matrices. For a detailed definition of the SVD technique, readers can refer to Zhao and Ye (2011).

4.1.2 PRCs' energy and energy entropy

When a fault occurs in a diesel engine, the corresponding energy distribution will change significantly. Therefore, the energy of PRCs, which can reflect the energy distribution in different frequency bands of the vibration signal, is used as the fault features in this study. In addition, to measure the randomness of energy distribution, the energy entropy is selected as the fault feature. Suppose $E_1, E_2, \ldots, E_n$ are the energy of the PRCs produced by CEITD. The energy entropy is defined as follows:

$$H_{\mathrm{EN}} = -\sum_{i=1}^{n} p_i \log p_i, \tag{22}$$

where $p_i=E_i/E$ represents the percentage of energy of each PRC_i in the signal energy, where the signal energy is $E=\sum_{i=1}^{n} E_i$.

4.1.3 Autoregressive model parameters

As a typical time-series analysis method, the autoregressive (AR) model can effectively reflect the essential characteristics of a dynamic system, and its coefficients are very sensitive to the variations in the condition (Cheng *et al.*, 2006). The AR model was established for each PRC_i as follows:

$$\mathrm{PRC}_i(t) = -\sum_{k=1}^{n} a_{ik}\mathrm{PRC}_i(t-k) + e_i(t), \tag{23}$$

where a_{ik} is the autoregressive coefficient, n is the model order, which is decided by the Akaike information criterion (AIC) (Shibata, 1976), and $e_i(t)$ is a Gaussian white noise series with zero mean and variance σ_i^2. Since the first several a_{ik} values contain the major characteristics of a dynamic system and the remnant variance σ_i^2 is closely related to its output characteristics (Cheng *et al.*, 2006), the first four autoregressive coefficients and the remnant variance of the PRCs were selected as the fault features.

4.2 Fault diagnosis approach for diesel engines based on CEITD and LSSVM

The fault diagnosis approach for diesel engines based on CEITD and LSSVM is as follows (Fig. 6):

Step 1: Collect the vibration signals of the diesel engine in normal and fault conditions, and divide the vibration signals into several samples.

Step 2: Each sample is decomposed into a series of PRCs and a residual signal by CEITD.

Step 3: Extract the singular values, PRCs energy and energy entropy, and AR model parameters of the first several PRCs as the fault feature vectors.

Step 4: Divide the samples into two subsets: the training samples and the testing samples.

Step 5: Use the fault feature vectors of the training samples to train the HDEPSO-LSSVM model.

Step 6: The fault feature vectors of testing samples are fed into the trained HDEPSO-LSSVM model, and the fault results are obtained.

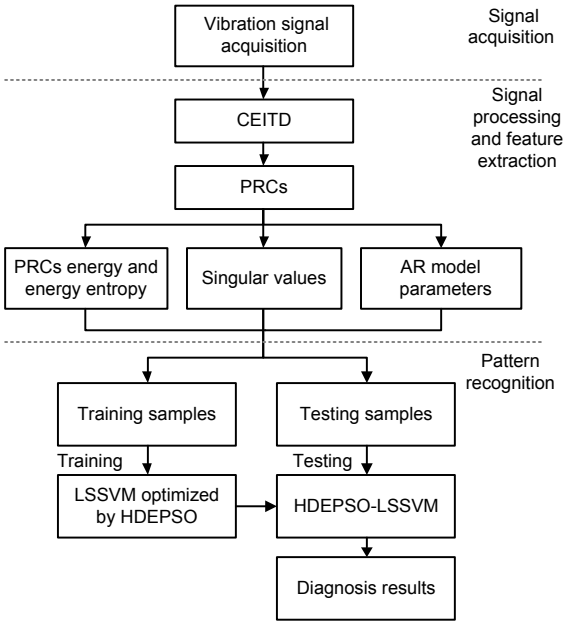

Fig. 6 Flowchart of the proposed fault diagnosis approach

5 Experiment and application

5.1 Experimental setup

All the data used in this work were collected from an experimental test rig (Fig. 7), which consists mainly of a diesel engine, a computer, and an LMS SCADA III multi-analyzer system (LMS Company, Belgium).

The 3D accelerometer is attached to the sixth cylinder head of the diesel engine, and the signals of its vertical direction were selected for fault diagnosis. The photoelectric sensor and the reflective stripe were used for the measurement of the engine speed and the top dead center position of the sixth cylinder. The LMS SCADA III multi-analyzer system was used to control the data acquisition, and the collected signals were stored in the computer for further analysis.

5.2 Fault simulation and data acquisition

One normal and five fault conditions were simulated on the sixth cylinder of the diesel engine; the specific parameters of the six working conditions are listed in Table 1. The clearance faults of intake and exhaust valve trains were simulated by adjusting the valve clearance adjustment screw of the sixth cylinder. Fuel supply faults were simulated by adjusting the fuel supply advance angle. Engine misfire refers to the phenomenon in which the mixture combustion flame cannot spread in the cylinder; i.e., there is no combustion in the cylinder. Therefore, misfire fault was simulated by cutting off the fuel supply system (Yang *et al.*, 2015; Zheng *et al.*, 2015). To reduce the crankshaft speed fluctuation and to keep the fault detection as simple as possible, the measurement was performed at the idle speed of 950 r/min. The sampling frequency was set to 25 000 samples per second, and a total of 30 samples under each condition were collected. Each sample contained 3158 data points, which correspond to one combustion cycle. Fig. 8 presents the vibration signals of the diesel engine in six working conditions.

Fig. 7 Experimental setup

1: 3D accelerometer; 2: photoelectric sensor; 3: reflective stripe

Table 1 Six working conditions of the diesel engine

Working condition	Description
1	Normal
2	Intake valve clearance increased by 0.2 mm
3	Exhaust valve clearance increased by 0.2 mm
4	Fuel supply advance angle decreased by 3°
5	Fuel supply advance angle increased by 3°
6	Misfire of the sixth cylinder

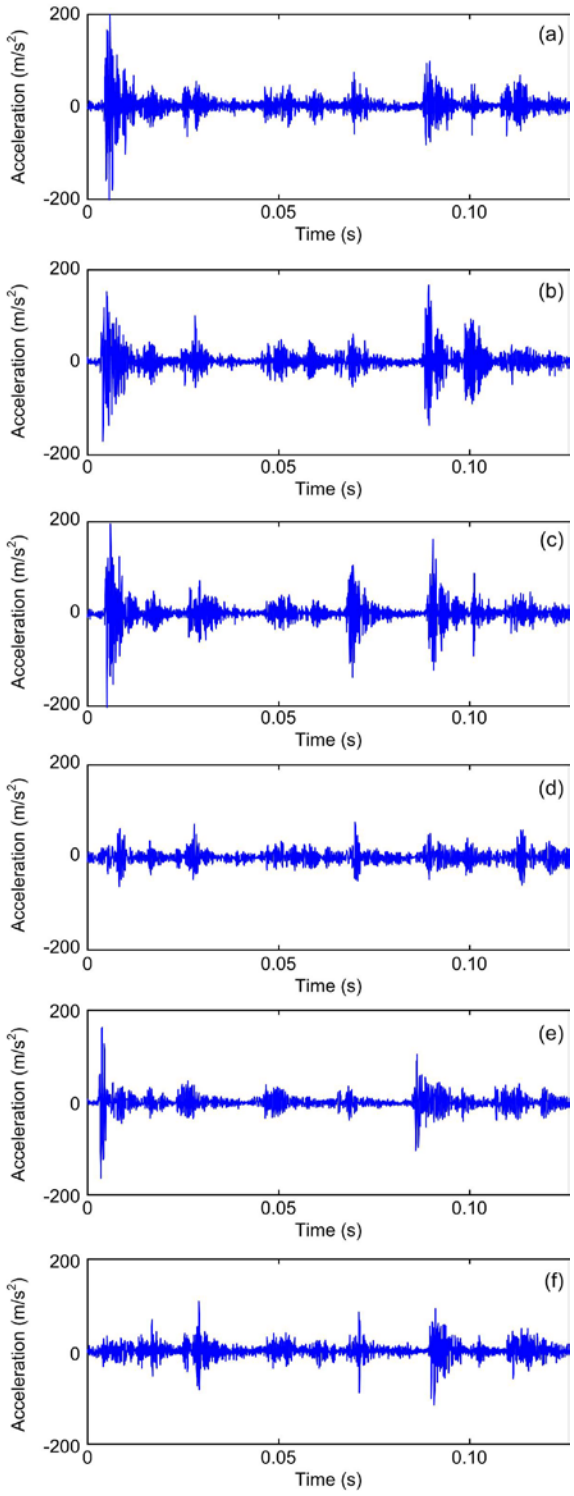

Fig. 8 Vibration signals of the diesel engine in six working conditions: (a) the normal condition; (b) the intake valve train in the excessive clearance condition; (c) the exhaust valve train in the excessive clearance condition; (d) the late fuel supply condition; (e) the early fuel supply condition; and (f) the 6th cylinder misfire condition

5.3 Application and comparison

5.3.1 Fault diagnosis results

The vibration signals of different working conditions were decomposed into several PRCs by using the CEITD method with the ensemble number 100 and the added noise amplitude 0.05 times the standard deviation of the signal. Considering the limited space, only the decomposition results of the early fuel supply condition (working condition 5) are presented in Fig. 9.

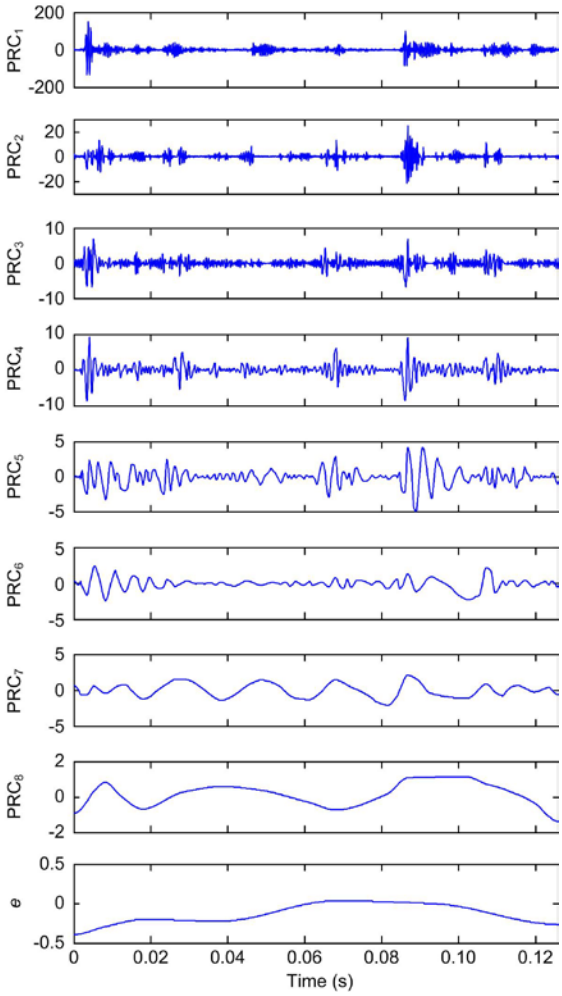

Fig. 9 CEITD decomposition results PRC_1, PRC_2, …, PRC_8 and the residue e of the diesel engine in the early fuel supply condition (Fig. 8e and working condition 5)

Since the information of the vibration signals of diesel engines is contained mainly in the first several PRCs, three typical types of time-frequency features, namely singular values, PRC energy and energy

entropy, and AR model parameters, were extracted from the first five PRCs and used as the fault feature vectors. For the purpose of eliminating the influence of the range of features, all the values of fault features were normalized into range [0, 1] and randomly divided into two subsets: the training set (20 samples of each condition) and the testing set (10 samples of each condition). Finally, the one-versus-one method was used to construct the multi-LSSVM model, and the proposed HDEPSO algorithm was used to optimize the regularization parameter γ and the kernel parameter σ of LSSVM. Here, the average classification accuracy of K-fold (K=5) cross-validation was selected as the fitness function. The searching range of γ and σ was set to (0, 100]. After implementing a number of different trials on the same processor, the maximum number of generations was set to 100, the number of individuals was set to 20 (10 for each direction), the scaling factor of mutation operation F was set to 0.9, the crossover probability CR was set to 0.5, the cognition learning factor c_1 was set to 1.5, the social learning factor c_2 was set to 1.7, the maximum inertia weight factor $w_{\max}$ set to 0.9, and the minimum inertia weight factor $w_{\min}$ was set to 0.4. The fitness curve of the HDEPSO algorithm is given in Fig. 10. After the optimization process was finished, the best γ was 65, and the best σ was 0.019. Training LSSVM with these parameters, the fault patterns of testing samples were obtained by putting the fault feature vectors into the trained HDEPSO-LSSVM classifier. The corresponding fault diagnosis results are listed in Table 2.

Table 2 Fault diagnosis results

Working condition	Recognition rate (%)					
	1	2	3	4	5	6
1	90	10	0	0	0	0
2	0	100	0	0	0	0
3	10	0	90	0	0	0
4	0	0	0	100	0	0
5	0	0	0	0	100	0
6	0	0	0	0	0	100

From Fig. 10, it can be observed that DE and PSO algorithms promote each other, and after 60 generations they converge to the overall optimal solution. From Table 2, it can be observed that only one sample of condition 1 is misclassified to condition 2,

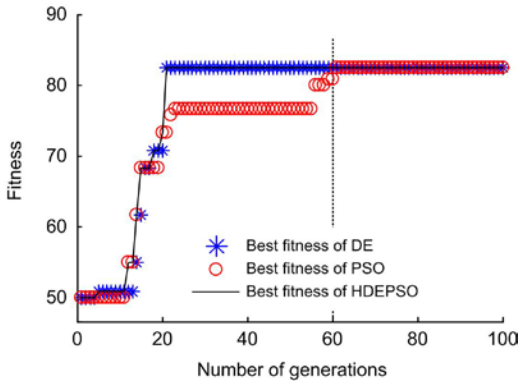

Fig. 10 Fitness curve of the HDEPSO algorithm

one sample of condition 3 is misclassified to condition 1, and the total recognition rate of the six working conditions is 96.3%. Therefore, it can be concluded that the proposed fault diagnosis approach can accurately identify the fault patterns of diesel engines.

5.3.2 Comparison

To further illustrate the effectiveness of the proposed fault diagnosis approach, several comparisons were made. First, the ITD, EEMD, and EITD methods were selected to replace the CEITD method in signal processing. The ensemble number and the added noise amplitude of EEMD and EITD were the same as those used for CEITD. Then, the singular values, energy and energy entropy, and AR model parameters of the first five components (PRCs or IMFs) were extracted as the fault feature vectors. Finally, to accomplish the fault diagnosis of the diesel engine, all fault features were normalized into the range 0 to 1, and the fault features obtained by each signal processing method were respectively fed into the LSSVM classifiers optimized by different parameter optimization methods, namely GA, PSO, DE, and HDEPSO. For GA, the maximum number of generations was set to 100, the number of individuals was set to 20, the mutation probability was set to 0.01, and the crossover probability was set to 0.4. The parameters of PSO, DE, and HDEPSO were the same as those used in Section 5.3.1. The fault diagnosis results are presented in Table 3.

From Table 3, it can be observed that the average fault diagnosis accuracy of using CEITD for signal processing is higher than that when using EEMD and EITD, and the average accuracy of using ITD is the lowest. The main reason is the appearance of mode-mixing in ITD, which makes some PRCs lose their

Table 3 Comparisons of different signal processing methods and parameter optimization methods

Signal processing method	Parameter optimization method	Recognition rate	Average recognition rate
ITD	GA	83.3%	83.3%
	PSO	80.0%	
	DE	83.3%	
	HDEPSO	86.7%	
EEMD	GA	91.7%	90.8%
	PSO	90.0%	
	DE	90.0%	
	HDEPSO	91.7%	
EITD	GA	91.7%	90.8%
	PSO	88.3%	
	DE	90.0%	
	HDEPSO	93.3%	
CEITD	GA	86.7%	92.0%
	PSO	91.7%	
	DE	93.3%	
	HDEPSO	96.7%	

physical meaning and provides false characteristic information. The mode-mixing problem can be solved to a certain extent by EEMD and EITD, so that the corresponding fault diagnosis accuracies of using EEMD and EITD for signal processing are higher than that when using ITD. However, besides mode-mixing alleviation, two problems were introduced. One is that the residual noise could not be completely canceled out; the other is that the final averaging was very difficult to implement. These two problems will cause large errors in the components produced by EEMD and EITD. In the CEITD method, the white noise is added in pairs (one positive and one negative), so that the residual noise can be almost completely eliminated. Moreover, the PRCs of CEITD are obtained in sequence by adding different amplitudes of the noise to the residue, so there is no difficulty in performing the final averaging. Therefore, the fault diagnosis accuracy of using CEITD for signal processing is the highest. In addition, it can be observed that no matter what signal processing method is used, the fault diagnosis accuracy of HDEPSO-LSSVM is higher than those of GA-LSSVM, PSO-LSSVM, and DE-LSSVM. This is due mainly to the fact that the information is transferred between DE and PSO in each generation during the optimizing process of the HDEPSO algorithm. By replacing the worst particle of PSO with the best individual of DE, or replacing the worst individual of DE with the best particle of PSO, HDEPSO realized advantageous complementarities between DE and PSO algorithms. Therefore, it is easier for the HDEPSO algorithm to jump out the local optimal solution and rapidly converge to the global optimal solution.

To show the efficacy of the fault features used in this study, the time-domain features, frequency-domain features, singular values, PRC energy and energy entropy, and AR model parameters were selected to replace the fault feature vectors used in Section 5.3.1. The time-domain features, including 11 commonly used time-domain statistical characteristics, were calculated for the first five PRCs, and a total of 55 time-domain features were obtained. The frequency-domain features, including 13 commonly used frequency-domain statistical characteristics, were also calculated for the first five PRCs and a total of 65 frequency domain features were obtained. The detailed definition of the time- and frequency-domain features can be found in Lei *et al.* (2007). Then, all fault features were normalized into the range 0 to 1, and HDEPSO-LSSVM was selected as the classifier for fault diagnosis. The fault diagnosis results with different fault feature vectors are shown in Table 4.

Table 4 Fault diagnosis results with different fault features

Fault feature	Recognition rate
Time-domain features	76.7%
Frequency-domain features	80.0%
Singular values	65.0%
Energy features	62.0%
AR model parameters	81.7%
Three typical time-frequency features	96.7%

Table 4 indicates that the time- and frequency-domain features, which have been successfully used in the fault diagnosis of rotating machinery, are inferior to the fault features used in the proposed approach. This is due mainly to the fact that the time- and frequency-domain features are not suitable for describing the characteristics of the transient and nonstationary vibration signals. Moreover, these features fail to reflect the relationship between different PRCs. By adopting singular values, PRC energy and energy entropy, and AR model parameters together as the fault feature vectors, the optimal fault diagnosis accuracy is obtained, and the combination

of different types of features is obviously superior to using one type of feature alone. The main reason is that the AR model can effectively reflect the essential characteristics of a dynamic system, and its coefficients are very sensitive to condition variation. Moreover, the singular values and PRC energy and energy entropy can be a supplement to the AR model parameters, although the fault diagnosis accuracy of using one of them alone as fault feature vectors is not high. Therefore, it can be concluded that the fault features used in this study are effective, and the combination of different types of time-frequency features is essential.

6 Conclusions

In this study, a novel approach based on CEITD and HDEPSO-LSSVM was proposed for the identification of diesel engine faults. First, targeting the nonstationary characteristics of the vibration signals of diesel engines, a new nonstationary signal processing method, ITD, was introduced to process the vibration signals. However, the mode-mixing problem, which makes the decomposition results devoid of physical meaning, often occurs when the ITD method is performed in practice. Therefore, to alleviate the mode-mixing problem of ITD, a CEITD method based on the noise-assisted analysis technique was presented. By applying the CEITD method to the vibration signals of the diesel engine, a set of stationary PRCs with no mode-mixing could be obtained. Second, the singular values, PRC energy and energy entropy, and AR model parameters of the first five PRCs were extracted as the fault feature vectors. Finally, targeting the parameter optimization problem of LSSVM, a HDEPSO algorithm was proposed, and the HDEPSO-LSSVM classifier was used to fulfill the fault diagnosis of the diesel engine. Experimental results showed that the proposed fault diagnosis approach could obtain satisfactory diagnosis results and offer superior fault diagnosis ability to traditional methods.

References

Ardia, D., Boudt, K., Carl, P., *et al.*, 2011. Differential evolution with DEoptim: an application to non-convex portfolio optimization. *R J.*, **3**(1):27-34.

Chen, B.J., He, Z.J., Chen, X.F., *et al.*, 2011. A demodulating approach based on local mean decomposition and its applications in mechanical fault diagnosis. *Meas. Sci. Technol.*, **22**(5):055704. http://dx.doi.org/10.1088/0957-0233/22/5/055704

Chen, M., Zheng, A.X., Jordan, M.I., *et al.*, 2004. Failure diagnosis using decision trees. Int. Conf. on Autonomic Computing, p.36-43. http://dx.doi.org/10.1109/ICAC.2004.31

Cheng, J.S., Yu, D.J., Yang, Y., 2006. A fault diagnosis approach for roller bearings based on EMD method and AR model. *Mech. Syst. Signal Process.*, **20**(2):350-362. http://dx.doi.org/10.1016/j.ymssp.2004.11.002

Cheng, J.S., Zheng, J.D., Yang, Y., 2012. A nonstationary signal analysis approach: the local characteristic-scale decomposition method. *J. Vibr. Eng.*, **25**(2):215-220 (in Chinese). http://dx.doi.org/10.3969/j.issn.1004-4523.2012.02.017

Cheng, M.Y., Hoang, N.D., Wu, Y.W., 2013. Hybrid intelligence approach based on LS-SVM and differential evolution for construction cost index estimation: a Taiwan case study. *Autom. Constr.*, **35**:306-313. http://dx.doi.org/10.1016/j.autcon.2013.05.018

Eberhart, R.C., Kennedy, J., 1995. A new optimizer using particle swarm theory. Proc. 6th Int. Symp. on Micro Machine and Human Science, p.39-43. http://dx.doi.org/10.1109/MHS.1995.494215

Frei, M.G., Osorio, I., 2007. Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non-stationary signals. *Proc. R. Soc. A*, **463**(2078): 321-342. http://dx.doi.org/10.1098/rspa.2006.1761

Hong, H., Wang, X.L., Tao, Z.Y., *et al.*, 2011. Centroid-based sifting for empirical mode decomposition. *J. Zhejiang Univ.-Sci. C (Comput. & Electron.)*, **12**(2):88-95. http://dx.doi.org/10.1631/jzus.C1000037

Huang, J., Hu, X., Geng, X., 2011. An intelligent fault diagnosis method of high voltage circuit breaker based on improved EMD energy entropy and multi-class support vector machine. *Electr. Power Syst. Res.*, **81**(2):400-407. http://dx.doi.org/10.1016/j.epsr.2010.10.029

Huang, N.E., Shen, Z., Long, S.R., *et al.*, 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A*, **454**(1971):903-995. http://dx.doi.org/10.1098/rspa.1998.0193

Huang, W., Kong, F., Zhao, X., 2015. Spur bevel gearbox fault diagnosis using wavelet packet transform and rough set theory. *J. Intell. Manuf.*, in press. http://dx.doi.org/10.1007/s10845-015-1174-x

Jiang, X., Li, S., Wang, Y., 2015. A novel method for self-adaptive feature extraction using scaling crossover characteristics of signals and combining with LS-SVM for multi-fault diagnosis of gearbox. *J. Vibroeng.*, **17**(4): 1861-1878.

Kadambe, S., Boudreaux-Bartels, G.F., 1992. A comparison of the existence of 'cross terms' in the Wigner distribution and the squared magnitude of the wavelet transform and the short-time Fourier transform. *IEEE Trans. Signal Process.*, **40**(10):2498-2517. http://dx.doi.org/10.1109/78.157292

Lei, Y.G., He, Z.J., Zi, Y.Y., *et al.*, 2007. Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. *Mech. Syst. Signal Process.*, **21**(5):2280-2294. http://dx.doi.org/10.1016/j.ymssp.2006.11.003

Lei, Y.G., He, Z.J., Zi, Y.Y., 2009. Application of the EEMD

method to rotor fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.*, **23**(4):1327-1338. http://dx.doi.org/10.1016/j.ymssp.2008.11.005

Li, J., Li, S., Chen, X., *et al.*, 2015. The hybrid KICA-GDA-LSSVM method research on rolling bearing fault feature extraction and classification. *Shock Vibr.*, **2015**:1-9. http://dx.doi.org/10.1155/2015/512163

Li, Y., Tse, P.W., Yang, X., *et al.*, 2010. EMD-based fault diagnosis for abnormal clearance between contacting components in a diesel engine. *Mech. Syst. Signal Process.*, **24**(1):193-210. http://dx.doi.org/10.1016/j.ymssp.2009.06.012

Li, Z., Yan, X., Yuan, C., *et al.*, 2011. Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. *Mech. Syst. Signal Process.*, **25**(7): 2589-2607. http://dx.doi.org/10.1016/j.ymssp.2011.02.017

Lin, J.S., 2012. Improved intrinsic time-scale decomposition method and its simulation. *Appl. Mech. Mater.*, **121-126**: 2045-2048. http://dx.doi.org/10.4028/www.scientific.net/AMM.121-126.2045

Mallipeddi, R., Suganthan, P.N., Pan, Q.K., *et al.*, 2011. Differential evolution algorithm with ensemble of parameters and mutation strategies. *Appl. Soft Comput.*, **11**(2): 1679-1696. http://dx.doi.org/10.1016/j.asoc.2010.04.024

Martin, W., Flandrin, P., 1985. Wigner-Ville spectral analysis of nonstationary processes. *IEEE Trans. Acoust. Speech Signal Process.*, **33**(6):1461-1470. http://dx.doi.org/10.1109/TASSP.1985.1164760

Martínez-Martínez, V., Gomez-Gil, F.J., Gomez-Gil, J., *et al.*, 2015. An artificial neural network based expert system fitted with genetic algorithms for detecting the status of several rotary components in agro-industrial machines using a single vibration signal. *Expert Syst. Appl.*, **42**(17-18):6433-6441. http://dx.doi.org/10.1016/j.eswa.2015.04.018

Moosavian, A., Ahmadi, H., Tabatabaeefar, A., *et al.*, 2013. Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing. *Shock Vibr.*, **20**(2):263-272. http://dx.doi.org/10.3233/SAV-2012-00742

Rilling, G., Flandrin, P., Gonçalvès, P., 2003. On empirical mode decomposition and its algorithms. IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, p.1-5.

Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrics*, **63**(1):117-126. http://dx.doi.org/10.1093/biomet/63.1.117

Storn, R., Price, K., 1997. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.*, **11**(4):341-359. http://dx.doi.org/10.1023/A:1008202821328

Su, Z., Tang, B., Liu, Z., *et al.*, 2015. Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. *Neurocomputing*, **157**:208-222. http://dx.doi.org/10.1016/j.neucom.2015.01.016

Suykens, J.A., Vandewalle, J., 1999. Multiclass least squares support vector machines. Int. Joint Conf. on Neural Networks, p.900-903. http://dx.doi.org/10.1109/IJCNN.1999.831072

Tay, F.E.H., Shen, L., 2003. Fault diagnosis based on rough set theory. *Eng. Appl. Artif. Intell.*, **16**(1):39-43. http://dx.doi.org/10.1016/S0952-1976(03)00022-8

Torres, M.E., Colominas, M., Schlotthauer, G., *et al.*, 2011. A complete ensemble empirical mode decomposition with adaptive noise. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, p.4144-4147. http://dx.doi.org/10.1109/ICASSP.2011.5947265

Vapnik, V.N., 1999. An overview of statistical learning theory. *IEEE Trans. Neur. Netw.*, **10**(5):988-999. http://dx.doi.org/10.1109/72.788640

Vong, C.M., Wong, P.K., 2011. Engine ignition signal diagnosis with wavelet packet transform and multi-class least squares support vector machines. *Expert Syst. Appl.*, **38**(7):8563-8570. http://dx.doi.org/10.1016/j.eswa.2011.01.058

Wang, C., Zhang, Y., Zhong, Z., 2008. Fault diagnosis for diesel valve trains based on time-frequency images. *Mech. Syst. Signal Process.*, **22**(8):1981-1993. http://dx.doi.org/10.1016/j.ymssp.2008.01.016

Wang, X., Liu, C., Bi, F., *et al.*, 2013. Fault diagnosis of diesel engine based on adaptive wavelet packets and EEMD-fractal dimension. *Mech. Syst. Signal Process.*, **41**(1): 581-597. http://dx.doi.org/10.1016/j.ymssp.2013.07.009

Wu, Z., Huang, N.E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data Anal.*, **1**(1):1-41. http://dx.doi.org/10.1142/S1793536909000047

Xie, Z., Shepard, W.S. Jr., Woodbury, K.A., 2009. Design optimization for vibration reduction of viscoelastic damped structures using genetic algorithms. *Shock Vibr.*, **16**(5):455-466. http://dx.doi.org/10.3233/SAV-2009-0480

Xu, H., Chen, G., 2013. An intelligent fault identification method of rolling bearings based on LSSVM optimized by improved PSO. *Mech. Syst. Signal Process.*, **35**(1-2): 167-175. http://dx.doi.org/10.1016/j.ymssp.2012.09.005

Xue, X., Zhou, J., Xu, Y., *et al.*, 2015. An adaptively fast ensemble empirical mode decomposition method and its applications to rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.*, **62-63**:444-459. http://dx.doi.org/10.1016/j.ymssp.2015.03.002

Yang, K., Ouyang, G., Li, A., *et al.*, 2015. Diesel engine misfire fault diagnosis based on instantaneous speed. Int. Conf. on Mechatronics, Electronic, Industrial and Control Engineering, p.1497-1501. http://dx.doi.org/10.2991/meic-15.2015.343

Zhang, X., Liang, Y., Zhou, J., *et al.*, 2015. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement*, **69**:164-179. http://dx.doi.org/10.1016/j.measurement.2015.03.017

Zhao, X., Ye, B., 2011. Selection of effective singular values using difference spectrum and its application to fault diagnosis of headstock. *Mech. Syst. Signal Process.*, **25**(5): 1617-1631. http://dx.doi.org/10.1016/j.ymssp.2011.01.003

Zheng, J.Y., Yang, Z.X., Wu, G.G., *et al.*, 2015. FTA-SVM-based fault recognition for vehicle engine. IEEE 12th Int. Conf. on Networking, Sensing and Control, p.180-184. http://dx.doi.org/10.1109/ICNSC.2015.7116031