

Optimizing SVM's parameters based on backtracking search optimization algorithm for gear fault diagnosis

Vantrung Thai¹, Junsheng Cheng², Viethung Nguyen³, Phuonganh Daothi⁴

^{1,2,3}State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha, China

^{1,3}Faculty of Mechanical Engineering, Hanoi University of Industry, Hanoi, Vietnam

⁴Faculty of Information Technology, University Economic and Technical Industries, Hanoi, Vietnam

¹Corresponding author

E-mail: ¹trongdhn@gmail.com, ²chenghnucn@hotmail.com, ³lenbaoluon@hotmail.com, ⁴daothi@phuonganh@gmail.com

Received 2 April 2018; received in revised form 29 July 2018; accepted 5 August 2018
DOI <https://doi.org/10.21595/jve.2018.19859>



Copyright © 2019 Vantrung Thai, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. The accuracy of a support vector machine (SVM) classifier is decided by the selection of optimal parameters for the SVM. The Backtracking Search Optimization Algorithm (BSA) is often applied to resolve the global optimization problem and adapted to optimize SVM parameters. In this research, a SVM parameter optimization method based on BSA (BSA-SVM) is proposed, and the BSA-SVM is applied to diagnose gear faults. Firstly, a gear vibration signal can be decomposed into several intrinsic scale components (ISCs) by means of the Local Characteristics-Scale Decomposition (LCD). Secondly, the MPE can extract the fault feature vectors from the first few ISCs. Thirdly, the fault feature vectors are taken as the input vectors of the BSA-SVM classifier. The analysis results of BSA-SVM classifier show that this method has higher accuracy than GA (Genetic Algorithm) or PSO (Particles Swarm Algorithm) algorithms combined with SVM. In short, the BSA-SVM based on the MPE-LCD is suitable to diagnose the state of health gear.

Keywords: signal processing, fault detection, gears, artificial neural networks, backtracking search optimization algorithm.

1. Introduction

The gearbox is the most crucial transmission mechanism in a rotating machine, and the implementation of online monitoring and diagnosis has become quite urgent and necessary. The vibration signal of a gearbox contains much noise and unstable factors thus the vibration signal is a typical non-stationary signal. The fault signal characteristic is very weak and is usually masked by noise, especially when the fault is in its early stages; thus, it is very difficult for getting the fault features [1]. There are three steps in diagnosing fault gears, namely: characteristic signal detection, feature extraction and fault classification. Generally, the process of extracting the features of faults is divided into selection and extraction. In both theoretical and experimental fields, the fact of selection and extraction of fault features is the common principle in diagnosing faults. Feature extraction is an important step in the fault identification. In case of incorrect or incomplete feature extraction, it unavoidably leads to the fact that the classification and false diagnosis are made with errors [2]. So, the extraction of the effective information about the fault characteristics from complex dynamic mechanical signals is the key factor to solve the problem of large-scale complex mechanical and electrical equipment fault diagnosis.

The corresponding characteristics of the vibration signals will be presented, and getting the extraction of the fault information of a non-stationary vibration signal is the crux of gear fault diagnosis. Vibration signals are usually processed by decomposing the original signal, such as fast Fourier transform, wavelet transform, Hilbert-Huang transform and EMD method [3]. This is one of the efficient techniques in diagnosing faults based on of time-frequency resolution, i.e. EMD is rather good at extracting the Intrinsic Mode Functions (IMFs), or mono-component functions

embracing the original signal [4]. But the EMD method has several shortcomings, such as the problem of mode mixing, distorted components and time-consuming decomposition [5]. To this end, Wu and Huang have proposed a method namely EEMD [6]. Pattern recognition is another side in diagnosing gear faults.

Recently, Cheng et al. have developed a new novel signal decomposition approach named as the local characteristic scale decomposition (LCD). Similar to the EMD, the LCD method is used to decompose a complex signal into several ISCs and a residue [7]. Because the LCD is the same EMD method, it is a kind of data-driven and adaptive non-stationary signal decomposition, so it is suitable for processing non-stationary signals such as vibration signals of gears.

Support vector machine (SVM) proposed by Vapnik et al. is a promising classification method and has been successfully applied to many engineering fields, such as face recognition, credit scoring and mechanical fault diagnosis, etc. [8]. SVM has the high generalization ability that depends on the adequate setting of parameters such as penalty coefficient C and kernel parameters σ . Therefore, the selection optimal parameters are essential and important to obtain a good performance in handling learning task of SVM [9]. In the past several years, the optimal SVM has gained great attentions. There are several methods for setting the parameters of SVM such as the grid search, the cross-validation method, and the gradient descent method [10]. These methods have several drawbacks; for example, the performance of grid search is sensitive to the setting of the grid range and coarseness for each parameter, which is not easy to be set without prior knowledge, and the cross-validation method requires long and complicated calculations.

The Backtracking Search Algorithm (BSA), which has been developed by Civicioglu, is an evolutionary algorithm (EA) to solve problems of real-valued numerical optimization. Selection, mutation and crossover are three well-known operators which are based on the BSA method but others like the genetic algorithm (GA) and differential evolution (DE) [11]. Furthermore, unlike many other metaheuristics, the BSA algorithm has only one control parameter and is not very sensitive to the initial value as reported therein. Since it had been introduced, the BSA attracted many researches, and it was applied to various optimization problems. Some successful examples are described below. In [12], a comparative analysis of BSA with other evolutionary algorithms for global continuous optimization was given. In [13], the BSA was used for the antenna array design. In [14], the BSA was employed for the design of robust Power System Stabilizers (PSSs) in multi machine power systems. In [15], it was used for the allocation of multi-type distributed generators along distribution networks [16]. The problems of real-valued numerical optimization were quickly resolved by the BSA method, and the experiments demonstrated better results than EAs. Because of applying only one control parameter, thereby, researchers got more advantages in conducting experiments. For analyzed positive point, the paper will apply BSA in the combination with the SVM aiming at diagnosing faults gears, so-called BSA-SVM.

In this paper, the MPE-LCD based on BSA-SVM is used to diagnose and classify gear faults. Firstly, the original vibration signal is decomposed by the LCD method into some ISC components. Then, the first few ISCs are chosen and extracted the fault feature vectors by the MPE algorithm [17]. Thirdly, in order to identify the work condition of gear, in this paper, the SVM is served as a classifier [18]. Furthermore, the authors proposed a new optimal algorithm named as BSA to make the optimization for the SVM parameters (BSA-SVM). The characteristic vectors of the stationary ISCs are taken as the input data of the BSA-SVM, and then the gear faults and the normal gear condition can be differentiated [7]. Furthermore, to ascertain the superiority of the MPE-LCD based on BSA-SVM, it is compared with EMD and EEMD combined with GA-SVM, PSO-SVM and SVM. The analysis results show that the diagnosis approach of BSA-SVM assisted with MPE-LCD is suitable for processing non-stationary signals, such as vibration signals of gears, with a higher accuracy and effect when compared to other methods.

The content of the paper composes as: Section 2 and 3 are dedicated to the LCD and MPE methods. In Section 4, the parameter optimization algorithm based on the BSA method is addressed. Section 5 refers to characteristic energy extracted from a number of ISCs which played as input vectors of BSA-SVM. The most important content will be shown in Section 6. This not

only introduces the optimization algorithm linked to BSOA-SVM and LCD but indicates its strong points in practice. Eventually, Section 7 will make some conclusions.

2. Local characteristics-scale decomposition method

Similar to EMD, the LCD is a self-adaptive data decomposition method. Based on the local characteristic scale, the LCD establishes a new definition of a mono-component with physical meaning [19]. Any complex signal $x(t)$ can be decomposed into several ISCs and the residue by using the LCD [18], as follows:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t), \quad (1)$$

where $c_i(t)$ is the i th ISC and $r_n(t)$ is the residue. The assessment criterion for ISC is as follows: all the maxima are positively, and all the minima are negatively in the whole data set. As shown in Fig. 1, any two adjacent maxima (minima), (t_{k-1}, x_{k-1}) and (t_{k+1}, x_{k+1}) are connected by a straight line. To the intermediate minima (maxima) (t_k, x_k) , its corresponding point (t_k, A_k) on this straight line is as follows:

$$A_k = x_{k-1} + \frac{t_k - t_{k-1}}{t_{k+1} - t_{k-1}}(x_{k+1} - x_{k-1}). \quad (2)$$

For assuring the smooth and symmetric features of the ISC, the proportions of x_k and A_k remain constant, as follows:

$$\frac{A_k}{x_k} = \frac{a - 1}{a}, \quad a \in (0,1), \quad (3)$$

where a is a proportional coefficient. When a is a constant, the proportions of x_k and A_k can remain constant. Generally, a is set as 0.5, thus, $A_k/x_k = -1$. At this time, x_k and A_k are symmetrical for the x -axis ensuring the symmetric feature of the achieved ISC.

Without losing the generality, the signal is shown as:

$$x(t) = x_1(t) + x_2(t), \quad (4)$$

where:

$$x_1(t) = [1 + 0.5 \sin(10\pi t)] \cos(100\pi t + 30\pi t^2),$$

$$x_2(t) = 4 \sin(40\pi t), \quad t = 0:1/1000:2.$$

In Fig. 1 the waveforms of $x(t)$, $x_1(t)$ and $x_2(t)$ are shown.

As a constant sequence or monotonic sequence, $r_n(t)$ represents the average trend of the signal generally. The ISC component c_i ($i = 1, \dots, n$) represents a signal of different frequencies from high to low. The signal frequency component in each frequency band changes according to the original signal. The decomposition process of the LCD is as follows (the whole decomposition process ends when the SD criterion [17] is satisfied, and a value of SD = 0.3 was set in this study):

(1) Assume the number of extrema of the signal $x(t)$ is M , and determine all the extrema (t_k, x_k) ($k = 1, \dots, M$) of the signal $x(t)$.

(2) Compute A_k ($k = 1, \dots, M$) according to Eq. (2), and calculate the corresponding L_k ($k = 1, \dots, M$) according to Eq. (3):

$$L_k = aA_k + (1 - a)x_k, \quad a \in (0,1). \quad (5)$$

where $a = 0.5$. When calculating A_1 , (t_0, x_0) needs to be extended by the image method [18]. Similarly, A_M can be calculated after (t_{M+1}, x_{M+1}) is extended [20].

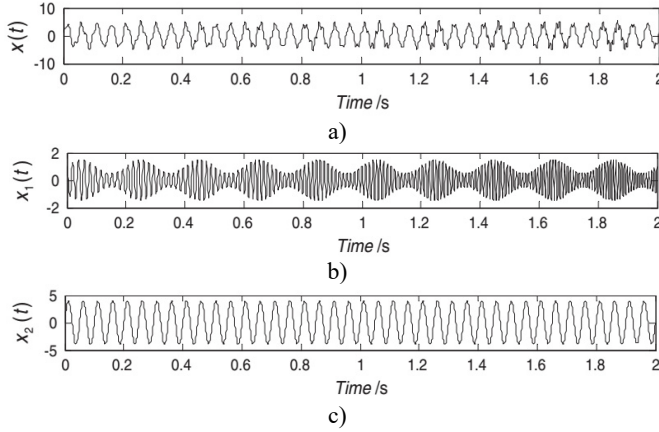


Fig. 1. Waveform of signal: a) $x(t)$, b) $x_1(t)$, c) $x_2(t)$

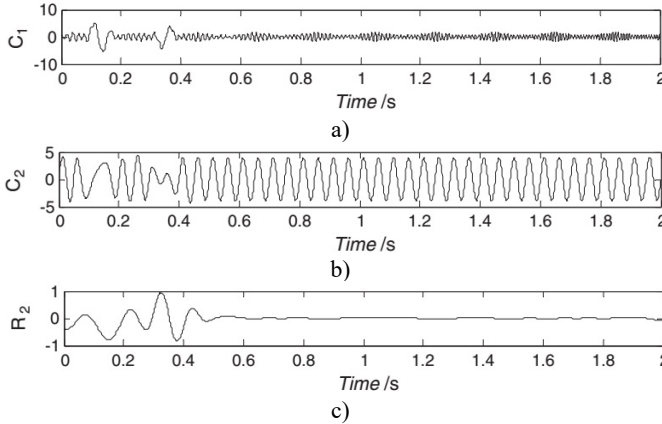


Fig. 2. Result of signal (10) decomposed by EMD

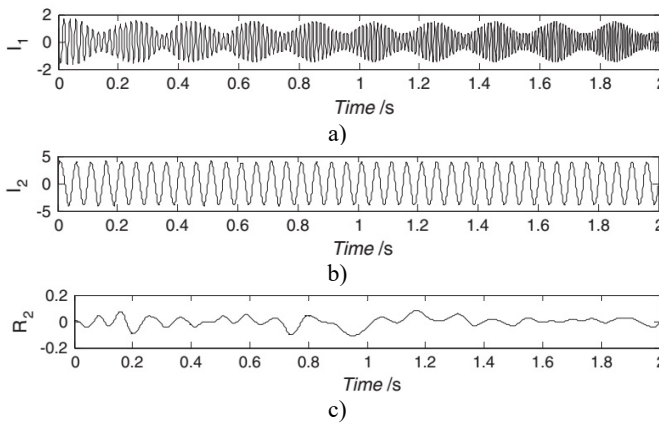


Fig. 3. Result of signal decomposed by LCD

(3) All the L_k ($k = 1, \dots, M$) are connected by a spline line $SL_1(t)$, which is defined as the

base line of the LCD. By contrast, EMD's base line is defined as the mean of the upper and the lower envelopes.

(4) The difference between the data and the base line $SL_1(t)$ is the first component $h_1(t)$:

$$h_1(t) = x(t) - SL_1(t). \quad (6)$$

If $h_1(t)$ is an ISC, take it as the first ISC of $x(t)$. Otherwise, take it as the original signal and repeat the above process until $h_m(t)$ is an ISC after m iterations of the computation. Afterwards, $h_m(t)$ is expressed as the ISC_1 .

(5) Separate the first ISC_1 from $x(t)$:

$$x(t) - ISC_1 = r_1. \quad (7)$$

(6) Take the residue r_1 as the original signal to be processed. Repeat the above process until ISC_1, \dots, ISC_n and the residues are obtained, as shown in Eq. (1).

The original signal is disintegrated into several ISCs by applying LCD, and the first few ISCs have the highest frequency and largest energy, while the last few ISCs remain relatively moderate. However, high frequency modulations still exist, which can interfere with the feature extraction. The TEO is used to demodulate the obtained ISC signal in this paper.

3. Permutation entropy (PE) and multi-scale permutation entropy (MPE)

3.1. Permutation entropy

Bandt and Pompe suggested the Permutation entropy (PE) in order to get the detection of the dynamic change in the time series by making a comparison between the neighboring values. The concept and steps to calculate PE values are described as follows [14]:

A time series given with the length $N\{x(k), k = 1, 2, \dots, N\}$, then the m dimensional vector at time i can be constructed as:

$$X_i^m = \{x(i), x(i + \tau), \dots, x[i + (m - 1)\tau]\}, \quad i = 1, 2, \dots, N - (m - 1)\tau, \quad (8)$$

where X_i^m is a new time series, m represents the embedding dimension, and τ is time delay. As described in [17], the X_i^m has a permutation $\pi_{r_0, r_1, \dots, r_{m-1}}$; if it satisfies that:

$$x(t + r_0\lambda) \leq x(t + r_1\lambda) \leq \dots \leq x(t + r_{m-1}\lambda), \quad (9)$$

where $0 \leq r_i \leq m - 1$ and $r_i \neq r_j$. Therefore, for an m -tuple vector, there are $m!$ possible distributions. Furthermore, we define the relative frequency for each distribution as:

$$P(i) = \frac{Num\{X_i^m\}}{N - (m - 1)\tau}, \quad (10)$$

where $Num\{X_i^m\}$ represents the number of X_i^m which is consistent with the type π . The definition of PE with m dimension can be written as:

$$H_{PE}(m) = \sum_i^{T-(m-1)\tau} p(i) \ln(p(i)). \quad (11)$$

Note that H_{PE} attains the maximum value $\ln(m!)$ when $P_i = 1/m!$. Then the normalized permutation entropy by $\ln(m!)$ could be formulated as:

$$H_{NPE}(m) = \frac{H_{PE}(m)}{\ln(m!)}. \quad (12)$$

It can be known that $H_{NPE}(m)$ satisfies that $0 \leq H_{NPE}(m) \leq 1$.

It would be concluded that a bigger H_{NPE} value shows the time series is more random and irregular. In case of a smaller H_{NPE} value, it can be deduced that the time series is more regular and periodic. In extreme cases, the time series is a white noise, the H_{NPE} value is one, also in the case of predicted signal (sine or cosine), the H_{NPE} value is zero. Therefore, PE may be utilized to make the estimation for the complexity as well as a dynamic change for a given signal.

3.2. Multi-scale permutation entropy

Costa [23] developed the multi-scale analysis algorithm to estimate the complexity of the original time series based on alternative scales. Aziz and Aric proposed the multi-scale permutation entropy (MPE) on the basis of multi-scale analysis definition. The MPE algorithm embraces two steps: firstly, getting multiple scale time series from the original time ones used the coarse-grained procedure, secondly, the determination of permutation entropy value for every coarse-grained time series [19]. This process is summed up as:

(1) A time series is given as $x(k)$, $k = 1, 2, \dots, N$. Splitting into disjointed windows is different in terms of length s . The coarse-grained time series y_j^s at a scale factor s in which s means a positive integer that may be constructed in accordance in Eq. (12):

$$y_j^s = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x_i, \quad 1 \leq j \leq \frac{N}{s}. \quad (13)$$

(2) In the MPE analysis, the PE of each coarse-grained time series is calculated based on Eqs. (7)-(11) and then plotted as the function of the scale factors, which can be expressed as:

$$MPE(x, s, m, \tau) = PE(y_j^s, m, \tau). \quad (14)$$

We calculated the PE of each coarse-grained time series as a function of scale factor t and then plotted it as a function of the scale factor, and call this procedure as the multi-scale permutation entropy. In order to select the best m for MPE calculation, we take the Gaussian white noise signal with length $N = 2048$ as an example. The MPEs are calculated under embedding dimension $m = 4, 5, 6$, and 7 when the parameters maximal scale factor $t_{max} = 12$ and $\tau = 1$ [7].

4. BSA and parameter optimization of SVM based on BSA

4.1. Backtracking search optimization algorithm

The main content of BSA method is to generate a trial population including two new crossover and mutation operators. The BSA strategy is considered as powerful exploration because this method both generates trial populations and controls the amplitude of the search-direction matrix and search-space boundaries [21]. Particularly, the BSA stores a population from a randomly chosen previous generation to use it in generating the search-direction matrix. The BSA strategy is very strong in terms of both a global exploration and local exploitation with a good feature of local minima avoidance [22, 23]. The main five procedures of the BSA method are presented as follows in Fig. 5.

Problem and Algorithm Parameter: An overview of five processes is provided as follows.

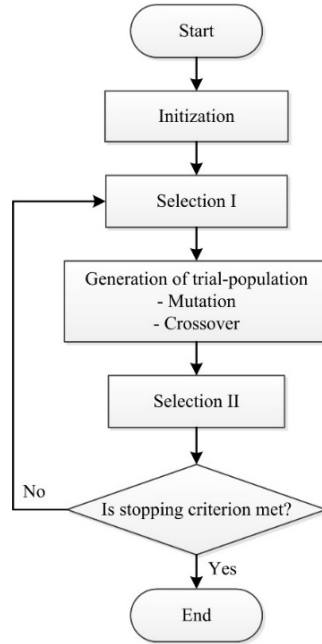


Fig. 4. Flow chart of backtracking search optimization algorithm

4.1.1. Initialization

The procedures of BSA begin from initializing the population P as follows:

$$P_{i,j} = lower_j + (upper_j - lower_j) \times random, i = 1,2, \dots, NP, j = (1,2, \dots, DP). \quad (15)$$

In which: NP is the size of population (Pop Size). DP is the dimensional number of the problem. random is a real value which is uniformly allocated between 0 and 1. $lower_j$ stands for the lower bound for the j th factor of the i th individual. $upper_j$ is the upper bound for the j th factor of the i th individual.

4.1.2. Selection-I

The generation of $oldP$ is the historical population, in the procedure of the Selection-I, and it is used to make a calculation for the search direction. The formulation is as follows:

$$P_{i,j} = lower_j + (upper_j - lower_j) \times random, i = 1,2, \dots, NP, j = (1,2, \dots, DP). \quad (16)$$

In each iteration, $oldP$ is defined as follows:

$$\text{if } a < b \text{ then } oldP := P|a, b \in [0,1]. \quad (17)$$

In which: $:=$ means the updated operation. Two random numbers, a and b , are within the range $[0, 1]$. The above equation assures that the BSA algorithm population might be randomly chosen from a historical population. The algorithm memorizes the historical population till it is changed via a random permutation.

4.1.3. Mutation

Initially, the trial population is created via mutation operation as per the bellow formulation:

$$T = P + (oldP - P) \times F, \quad (18)$$

where F is a scale factor which controls the amplitude of mutation ($oldP - P$).

In this paper, $F = 3 \cdot random$, where $random$ is a random real number with uniform distribution within the range (0, 1). By involving the historical population in calculating $oldP - P$, BSA learns from its memory of previous generations to obtain a trial population [24].

4.1.4. Crossover

The final trial population T is generated by crossover. The guidance of the trial individuals characterized by improved fitness values supports the search direction in optimizing the problem. BSA crossover works as the following procedure. A binary integer-valued matrix (map) of size $NP \times DP$ is computed in the first step.

The individuals of T are generated by utilizing relevant individuals of P . If $map_{i,j} = 1$, T is updated with $T_{i,j} = P_{i,j}P_{i,j}$.

4.1.5. Selection-II

In the stage of Selection-II, T_i makes the corresponding P_i better in the manner of fitness value which is utilized for updating P_i [21]. As the BSA finds the best optimal value (P_{best}) which is dominant comparing to all previous ones, P_{best} takes the place of all optimal solutions, and these values are updated as the most fitting value of P_{best} .

4.2. SVM parameter optimization based on BSA

4.2.1. Support vector machine (SVM)

The SVM is a type of machine learning technique. The SVM relies on the theory of statistical learning. The SVM is handling the training samples as the input to a higher-dimensional characteristic space through the use of a mapping function ϕ^7 . Assuming that there is a given set of the training samples $G = \{(x_i, y_i), i = 1, 2, \dots, l\}$ in which each sample $x_i \in R^d$ belongs to a class by $y \in \{+1, -1\}$, and the training data is not linearly separable in the space of feature, then the target function can be expressed as follows [17]:

$$\text{Minimize } \phi(\omega) = \frac{1}{2} \langle \omega, \omega \rangle + C \sum_{i=1}^l \xi_i, \quad y_i(\langle \omega, \phi(x_i) \rangle + b) \geq 1 - \xi_i, \quad (19)$$

$$\xi_i \geq 0, \quad i = \{1, 2, \dots, l\}.$$

In which ω is the normal vector of the hyperplane, C is the penalty parameter, b is the bias, ξ_i is non-negative slack variables, and $\phi(x)$ is the mapping function.

By introducing a set of Lagrange multipliers $\alpha_i \geq 0$, the optimization problem could be rewritten as:

$$\text{Maximize } L(\omega, b, \alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i x_j), \quad (20)$$

$$0 \leq \alpha_i \leq C, \quad \sum_{i=1}^l \alpha_i y_i = 0.$$

The function of making decision can be achieved as:

$$f(x) = \text{sgn} \left[\sum_{i=1}^l \alpha_i y_i K(x, x_i) + b \right]. \quad (21)$$

The SVM method used the radial basis kernel function which is the most common kernel function as it indicates in the bellow equation:

$$K(x, x_i) = \exp \left(-\frac{\|x - x_i\|^2}{2\sigma^2} \right), \quad (22)$$

where σ is the kernel parameter.

4.2.2. SVM's parameter optimization relied on BSA

The performance of SVM is significantly impacted by its parameters. It needs to be selected as a penalty factor C and kernel parameter σ in the function of Gaussian kernel. It is not easy to select these parameters. As a whole, C and σ are chosen empirically. Therefore, this paper applied the BSA as a technique for optimizing the parameters of SVM. As a consequence, C and σ played as the variables. SVM error testing became the fitness function for the sake of optimization.

The SVM error testing is made as:

$$\text{fitness}(x) = \text{TestError}_{SVM}(x), \quad (23)$$

where $x = (C; \sigma)$ and the test error of SVM was defined as:

$$\text{TestError}_{SVM} = \left(1 - \frac{y_t}{y_t + y_f} \right) 100 \%, \quad (24)$$

where y_t and y_f mean the numbers of true and false classified samples, respectively. The desirable value is too small to ensure high classification accuracy.

4.2.3. Experimental results

The paper used Thyroid, Seed and Escherichia coli (E. Coli) as three common sets of benchmark data provided by the University of California, Irvine to quantify the performance of the suggested BSA-SVM method.

Table 1 showed the training and test sets. Every sampling set is separated into two subsets. In which, one subset is used for SVM training, and the other is for testing the obtained model. The training set accounted for 70 %, and the test set took 30 % of the total samples. Based on the trial and error detection practice, this proportion was selected in order to evaluate the performance of the achieved SVM which was optimal for the available samples.

Three data sets were used to make classification among the BSA-SVM, GA-SVM, PSO-SVM, and CMAES-SVM methods. The values selection is the same in all four above methods in order to make comparison fairly (e.g., iteration = 30 and Pop Size = 30). For the PSO, the parameters were fixed with the values given in the literature [22, 26] (i.e., $W = 0.9$, $c_1 = 0.5$, and $c_2 = 1.25$). For CMAES, the parameters were fixed with the values given in the literature [24] (i.e., $\sigma_0 = 0.25$ and $\mu = 4 + 3\log(N)$). Each testing method result was considered as the average value of 30 runs. The training data and test data were both mixed and were randomly divided, as shown in Table 1.

Lin et al. argued that [25] the lower and upper bounds of C were given in [0:01; 35000] and σ in [0:01; 32] for the BSA-SVM, GA-SVM, PSO-SVM, and CMAES-SVM classifiers. Each search method gave the values of C and σ in order to give the smallest value of the classification

error. Table 2 presented experimental results.

The effectiveness of the proposed method was illustrated in Tables from 2 to 4 on the basis of the detailed classification results of each data set. The Thyroid and Seed data sets encompassed three classes, thus, the authors needed two SVM classifiers. The E. Coli data set embraced five classes; therefore, the authors needed four classifiers. These tables show the optimal parameters (C and σ), average test error, and average cost time done by different algorithms.

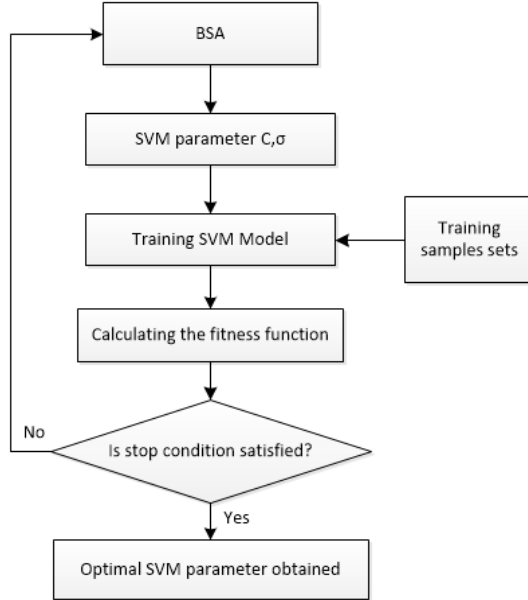


Fig. 5. Parameter optimization flowchart of SVM based on BSA

Table 1. Parameters of datasets

Name	Data	Train	Test	Validation data	Input	Class
Thyroid	215	151	52	150	5	3
Seed	210	147	55	160	7	3
E. coli	327	229	60	180	7	5

Table 2. Results of thyroid data set

Method	Training samples	Test samples	Average cost time (s)	Average test errors (%)	Refs.
BSA-SVM	150	52	15.83	0.000	[25]
PSO-SVM	150	52	36.31	2.358	[26]
GA-SVM	150	52	31.12	1.442	[27]
CMAES-SVM	150	52	32.56	2.622	[28]

Table 3. Results of seed data set

Method	Training samples	Test samples	Average cost time (s)	Average test errors (%)	Refs.
BSA-SVM	160	55	17.65	0.910	[25]
PSO-SVM	160	55	40.69	3.993	[26]
GA-SVM	160	55	36.38	3.552	[27]
CMAES-SVM	160	55	37.56	2.298	[28]

The results of the E. Coli, Seed and Thyroid classification were given in Table 1. The next Table indicates that the cost timing and testing error of BSA-SVM was a bit lower in comparison with those of other methods as GA-SVM, PSO-SVM, and CMAES-SVM. Civicioglu said that the BSA utilized a mutation mechanism in which there was a complex crossover and one individual [25]. Moreover, by using its memory, the BSA took an advantage of the experiences achieved

from previous generations. Tables 2-4 demonstrate that the BSA-SVM classifier achieved higher classification accuracy in the way of a shorter time as comparing to other methods. Next, the BSA-SVM method was used to diagnose a gear fault.

Table 4. Results of E. Coli data set

Method	Training samples	Test samples	Average cost time (s)	Average test errors (%)	Refs.
BSA-SVM	180	60	26.73	4.641	[25]
PSO-SVM	180	60	59.43	5.364	[26]
GA-SVM	180	60	44.78	7.787	[27]
CMAES-SVM	180	60	42.52	8.582	[28]

5. Gear fault diagnosis method based on BSA-SVM and LCD-MPE

Sections 2 and 3 showed the applied LCD in the combination with the MPE under different work status to analyze a gear vibration signal. Thereby, fault patterns are clearly recognized via the energy of each ISC changed when the gear malfunctioned. The BSA-SVM input vector is the energy feature of every ISC component. By doing so, the work status and fault patterns of the gear can be identified.

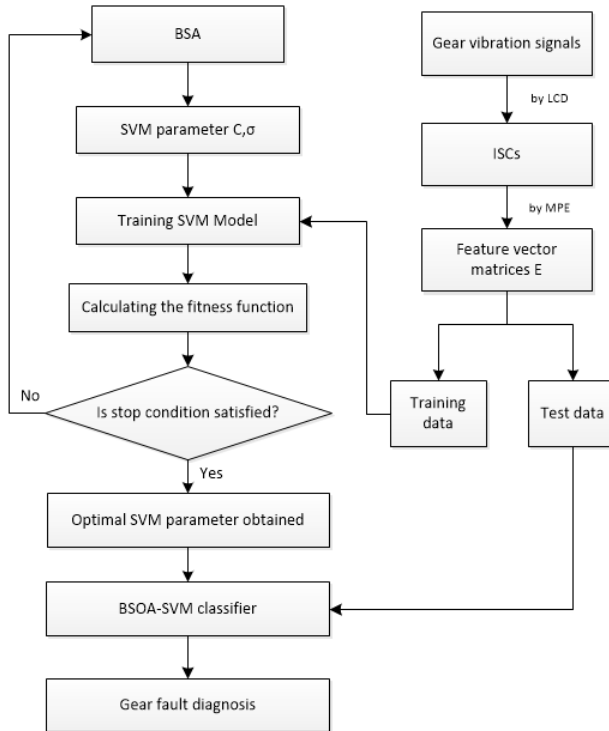


Fig. 6. Flowchart of gear fault diagnosis based on LCD-MPE and BSA-SVM

Fig. 6 shows the flow chart of the LCD-MPE and BSA-SVM method for the gear fault diagnosis. The description of the method for diagnosing gear faults is as follows:

(1) Choose several signals as samples under three cases: gear normal, gear chipped, and gear broken.

(2) Disintegrate the original vibration signals by the LCD method into several ISCs. The first m ISCs which include the most dominantly information of the fault gear will be chosen to extract the feature.

(3) Calculate MPE for each gear vibration signal getting from the first m ISCs with the

parameter selection: $m = 6$; $\tau = 1$; $N = 2048$, and the maximum scale factor $t_{max} = 12$.

(4) Then the MPEs are obtained in all scales, and are viewed as the feature vector E to represent the main fault information of gear vibration signal.

(5) Run the BSA-SVM, with the parameters of SVM, named as N and σ , which are optimized by the BSA.

(6) Input the vector E into the BSA-SVM classifier and realize training. The fitness function is given by Eq. (21). The net results of N and are input to the BSA-SVM classifier.

(7) The output of BSA-SVM is defined with the work conditions of the gear with three classes: class 1 – normal $[-1 -1 +1]$; class 2 – chipped tooth $[-1 +1 -1]$; class 3 – broken tooth $[+1 -1 -1]$.

6. Application of BSA-SVM and LCD-MPE to diagnose gear faults

Data collected in this paper comes from the public data sets distributed by the Prognostics and Health Management (PHM) Society [29], and the data sets were sampled synchronously from two accelerometers mounted on each side of the gearbox housing shown in Fig. 7. Data collection was implemented at five different shaft speeds of 30, 35, 40, 45 and 50 Hz under a low and high load from the brake with the sample frequency $fs = 200/3$ kHz. The low and high loads have resulted in a small difference of shaft speed, each load case is repeated four times. There are totally 560 different cases to be diagnosed, with six cases using helical gears and the other eight spur gears. On the input shaft, the tachometer generates 10 pulses per revolution and data from the tachometer is very accurate. Here we chose the helical gear to research with three states: normal, chipped tooth and broken tooth, and the sampling frequency can be taken as 1024 Hz. The vibration signals of helical gear with three conditions (normal, chipped tooth and broken tooth) were tested, and 40 vibration signals from the helical gear in each condition were taken from 6 groups collected at random as the test data.

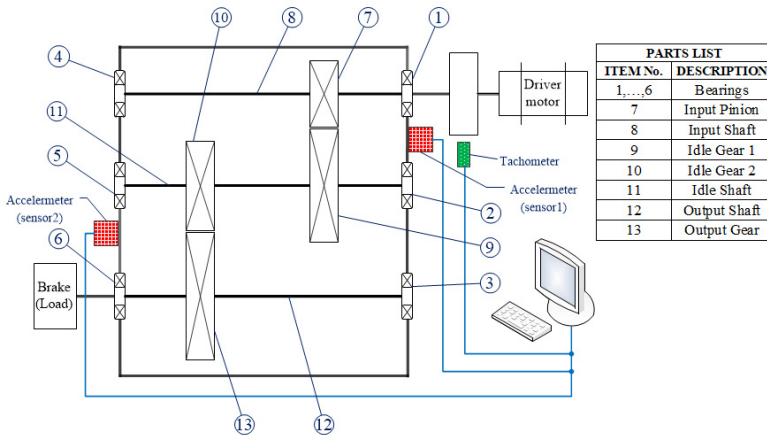


Fig. 7. Common test rig

Firstly, the vibration signals of each group with three conditions named as normal, chipped tooth and broken tooth were decomposed by LCD-MPE into a number of ISCs. The first seven ISCs embrace the most dominant information which is selected and arranged in accordance with the frequency components as $c_1(t), c_2(t), \dots, c_8(t)$ from high to low. Then, the fault features vector T was obtained following the MPE method. Finally, the BSA-SVM was used to identify the various patterns. In order to have a fair comparison, the original vibration signals are chosen similarly. These signals are decomposed by LCD into ISCs, of which, the first eight IMFs are selected and arranged in the manner of high to low values basing on frequency components as $c_1(t), c_2(t), \dots, c_8(t)$. And then, the feature vector is taken into the input of BSA-SVM. The identification results for the test samples based on LCD-MPE and compared with EMD and

EEMD method are shown in Table 5.

The input vector of the BSA-SVM method uses the LCD or EMD analysis as preprocessor to extract the energy in each frequency band in identifying the fault gears. These input vectors affect the identification of the fault gears. The frequency components after decomposition by LCD-MPE has better results than the ones in the EMD and EEMD method, therefore, the method BSA-SVM based on LCD is better than that based on EMD. The experimental results were demonstrated in Table 5 and Fig. 8. As a result, the accuracy of identification is good, and the time is shorter. In practical cases, they may be used for a fault diagnosis.

Table 5. Gear fault diagnosis results based on LCD-MPE or EEMD-MPE or EMD-MPE

Signal	Method	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	Test error
S ₁	LCD	0.6841	0.5156	0.4611	0.4267	0.2591	0.1648	0.0772	0.0239	Normal (0 %)
	EMD	0.6235	0.4649	0.4299	0.4010	0.1801	0.1252	0.0369	0.0127	Normal (2.18 %)
	EEMD	0.6518	0.5233	0.3836	0.2925	0.2451	0.0849	0.0327	0.0059	Normal (1.09 %)
S ₂	LCD	0.7329	0.5182	0.4869	0.4218	0.2234	0.1152	0.0993	0.0072	Chipped tooth (0 %)
	EMD	0.6946	0.4619	0.4461	0.3967	0.1846	0.0946	0.0856	0.0061	Chipped tooth (2.18 %)
	EEMD	0.7358	0.5821	0.4812	0.3519	0.2183	0.0612	0.0319	0.0052	Chipped tooth (1.09 %)
S ₃	LCD	0.7365	0.5285	0.4965	0.4378	0.1958	0.0983	0.0941	0.0041	Broken tooth (0 %)
	EMD	0.6562	0.4925	0.4521	0.3876	0.1732	0.0729	0.0873	0.0037	Broken tooth (2.18 %)
	EEMD	0.6832	0.4719	0.4028	0.4001	0.1296	0.0652	0.0372	0.0062	Broken tooth (1.09 %)

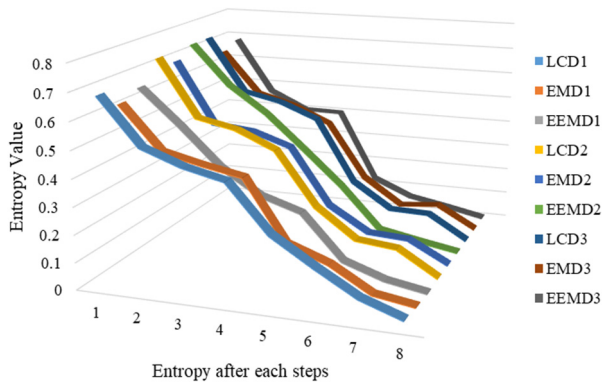


Fig. 8. Results based on LCD-MPE or EEMD for gear fault diagnosis

7. Conclusions

Basing on the feature of non-stationary of gear fault signals, a method of diagnosing a faulty gear based on the LCD-MPE and BSA-SVM is proposed. LCD and MPE were firstly used to pre-process different types of vibration signals. BSA was then employed to make the optimization for the parameters of SVM aiming at increasing the sensitivity and accuracy of SVM. This combination used to classify data of the work condition of the gear is called as BSA-SVM. When the working condition of gear changes, then the fault feature extraction is also changed indicating that the energy of each frequency component is changed when the gear with different faults is operating. Thus, the feature vector of each ISC component is adopted as input features for BSA-SVM to identify the gear work condition. Based on experimental results, some conclusions of the paper are:

(1) LCD is a method of processing self-adaptive signals, which can be applied to nonlinear and non-stationary processes faultlessly.

(2) BSA is an optimal algorithm that is used to make the optimization for the parameters of SVM network. Experimental results above show that the BSA and SVM combine better than other optimization algorithms (GA, PSO) combined with SVM.

(3) The successful combination of LCD-MPE with BSA-SVM identifies the work condition and fault patterns for the gears and provides an effective tool for intelligent fault diagnosis of gears.

(4) The BSA-SVM method that took the fault feature vector of each frequency component based on LCD combined with MPE as the input features has greater identification power than that based on EMD or EEMD combined with MPE.

In short, the paper presents another approach to gear fault diagnosis using a chain of data mining methods. Firstly, the LCD is used for vibration signals decomposition. The LCD is an improved form of the Huang-Hilbert Transform. Thanks to this, the problem of the mode mixing in this transform is removed. It is well visible in diagnosis results presented in Table 5. Next, the MPE is calculated as a feature characterizing each ISC obtained after signal decomposition by the LCD method. Then the MPEs values obtained in all scales are used as the feature vector to SVM classifier. However, the paper authors have improved the effectiveness of this classifier by optimizing its parameters with the aid of backtracking search optimization algorithm (BSA). Thanks to such a novel approach to the whole procedure of data mining applied to gear fault diagnosis they propose a very effective methodology which is able to perform the bearing different fault type diagnosis in a very reliable and robust way. According to the results presented in Table 5, the combination of LCD-MPE with BSOA-SVM can identify the work condition and fault patterns of the gears with the test error equal to 0 % for all the tested cases.

Acknowledgements

The authors first would like to thank the financial support by the Chinese National Science Foundation Grants (No. 51375152 and 51175158). The authors also would like to thank the support from the Collaborative Innovation Center of Intelligent New Energy Vehicle, and the Hunan Collaborative Innovation Center for Green Cars.

References

- [1] Cui L., Yao T., Zhang Y., Gong X., Kang C. Application of pattern recognition in gear faults based on the matching pursuit of a characteristic waveform. *Measurement*, Vol. 104, 2017, p. 212-222.
- [2] Zheng Z., Jiang W., Wang Z., Zhu Y., Yang K. Gear fault diagnosis method based on local mean decomposition and generalized morphological fractal dimensions. *Mechanism and Machine Theory*, Vol. 91, 2015, p. 151-167.
- [3] Ao H., Cheng J., Yang Y., Truong T. K. Support vector machine parameter optimization method based on artificial chemical reaction optimization algorithm and its application to roller bearing fault diagnosis. *Journal of Vibration and Control*, Vol. 21, Issue 12, 2013, p. 2434-2445.
- [4] Luo S., Cheng J., Ao H. Application of LCD-SVD technique and CRO-SVM method to fault diagnosis for roller bearing. *Shock and Vibration*, Vol. 1, 2015, p. 2015-8.
- [5] Cheng J., Yu Tang D. J., Yang Y. Local rub-impact fault diagnosis of the rotor systems based on EMD. *Mechanism and Machine Theory*, Vol. 44, Issue 4, 2009, p. 784-791.
- [6] Le D., Cheng J., Yang Y., Tran T., Pham V. Gears fault diagnosis method using ensemble empirical mode decomposition energy entropy assisted ACROA-RBF neural network. *Journal of Computational and Theoretical Nanoscience*, Vol. 13, Issue 5, 2016, p. 3222-3232.
- [7] Le D., Cheng J., Yang Y., Pham M., Thai V. Gear fault diagnosis method based on local characteristic-scale decomposition multi-scale permutation entropy and radial basis function network. *Journal of Computational and Theoretical Nanoscience*, Vol. 14, Issue 10, 2017, p. 5054-5063.
- [8] Zhang X., Qiu D., Chen F. Support vector machine with parameter optimization by a novel hybrid method and its application to fault diagnosis. *Neurocomputing*, Vol. 149, 2015, p. 641-651.
- [9] Bordoloi D. J., Tiwari R. Optimum multi-fault classification of gears with integration of evolutionary and SVM algorithms. *Mechanism and Machine Theory*, Vol. 73, 2014, p. 49-60.

- [10] **Saravanan N., Siddabattuni V. N. S. K., Ramachandran K. I.** Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM). *Applied Soft Computing*, Vol. 10, Issue 1, 2010, p. 344-360.
- [11] **Chen D., Lu R., Zou F., Li S., Wang P.** Learning and niching based backtracking search optimisation algorithm and its applications in global optimisation and ANN training. *Neurocomputing*, Vol. 266, 2017, p. 579-594.
- [12] **Mandal S. R. S., Mittal K.** Comparative analysis of backtrack search optimization algorithm (BSA) with other evolutionary algorithms for global continuous optimization. *International Journal of Computer Science and Information Technology*, Vol. 6, 2015, p. 3237-41.
- [13] **Civicioglu P.** Circular antenna array design by using evolutionary search algorithms. *Progress in Electromagnetics Research*, Vol. 54, 2013, p. 265-84.
- [14] **Shafullah A. M. M., Coelho L. S.** Design of robust PSS in multimachine power systems using backtracking search algorithm. *18th International Conference on Intelligent System Application to Power Systems*, 2015.
- [15] **Attia El-Fergany** Multi-objective allocation of multi-type distributed generators along distribution networks using backtracking search algorithm and fuzzy expert rules. *Electric Power Components and Systems*, Vol. 44, Issue 2016, 3, p. 252-267.
- [16] **Chaib A. E., Boucekara H. R. E. H., Mehasni R., Abido M. A.** Optimal power flow with emission and non-smooth cost functions using backtracking search optimization algorithm. *International Journal of Electrical Power and Energy Systems*, Vol. 81, 2016, p. 64-77.
- [17] **Li Y., Xu M., Wei Y., Huang W.** New rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement*, Vol. 77, 2016, p. 80-94.
- [18] **Zheng J., Cheng J., Yang Y.** Rolling bearing fault diagnosis approach based on LCD and fuzzy entropy. *Mechanism and Machine Theory*, Vol. 70, 2013, p. 441-453.
- [19] **Liu H., Wang X., Lu C.** Rolling bearing fault diagnosis based on LCD-TEO and multifractal detrended fluctuation analysis. *Mechanical Systems and Signal Processing*, Vol. 60, Issue 61, 2015, p. 273-288.
- [20] **Deng W. W. Y., Qian C., Wang Z., Dai D.** Boundary-processing-technique in EMD method and Hilbert transform. *Chinese Science Bulletin*, Vol. 46, 2001, p. 954-960.
- [21] **Costa A. L. G. M., Peng C. K.** Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, Vol. 92, 2002, p. 68102.
- [22] **Ebbesen S., Kiwitz, Guzzella L.** Generic particle swarm optimization Matlab function. *Proceedings of the American Control Conference*, 1519, p. 1524-2012.
- [23] **Wang Y., Cai, Zhang Q.** Differential evolution with composite trial vector generation strategies and control parameters. *IEEE Transactions on Evolutionary Computation*, Vol. 15, 2011, p. 55-66.
- [24] **Lin S. W., Ying K. C., Chen S. C., Lee Z. J.** Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, Vol. 35, 2008, p. 1817-1824.
- [25] **Civicioglu P.** Backtracking search optimization algorithm for numerical optimization problems. *Applied Mathematics and Computation*, Vol. 219, Issue 15, 2013, p. 8121-8144.
- [26] **Cervantes J., Garcia Lamont F., Rodriguez L., López A., Castilla J. R., Trueba A.** PSO-based method for SVM classification on skewed data sets. *Neurocomputing*, Vol. 228, 2017, p. 187-197.
- [27] **Huang Y., Wu D., Zhang Z., Chen H., Chen S.** EMD-based pulsed TIG welding process porosity defect detection and defect diagnosis using GA-SVM. *Journal of Materials Processing Technology*, Vol. 239, 2017, p. 92-102.
- [28] **Jafrasteh B., Fathianpour N.** Hybrid simultaneous perturbation artificial bee colony and back-propagation algorithm for training a local linear radial basis neural network on ore grade estimation. *Neurocomputing*, Vol. 235, 2017, p. 217-227.
- [29] **Society P.** Data Analysis Competition. <http://www.phmsociety.org/competition/PHM/09/apparatus>, 2009.



Van Trong Thai received a Master degree in mechanical engineering from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2011. Now he works at the Faculty of Mechanical Engineering, Hanoi University of Industry, Hanoi, Vietnam. His current research on fault diagnosis for achieving a Ph.D. degree at the Hunan University, Changsha, China.



Jun Sheng Cheng is a Ph.D. Professor in the Hunan University, Changsha, China. Now he works at the College of Mechanical and Vehicle Engineering, State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University. His current research interests include vibration and noise, signal processing and fault diagnosis.



Viet Hung Nguyen received a Master degree in mechanical engineering from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2008. Now he works at the Faculty of Mechanical Engineering, Hanoi University of Industry, Hanoi, Vietnam. His current research on fault diagnosis for achieving a Ph.D. degree at the Hunan University, Changsha, China.



Phuong Anh Dao Thi received a Master degree in information systems from the Military Technical Academy, Hanoi, Vietnam, in 2010. Now she works at the Faculty of Information Technology, University Economic and Technical Industries, Hanoi, Vietnam. Her current research interests include information system and neural network.