Fault diagnosis of gearbox based on ant colony algorithm optimized support vector machine

Shilin Sun¹,*, Renxiang Lu²,*

¹Business School of Shanghai Dianji University, Shanghai, 201306, China
²Business School of Shanghai Dianji University, Shanghai, 201306, China

*Corresponding author’s e-mail: lrx524@sina.com

Abstract. The kernel function parameter g and penalty coefficient c in Support Vector Machine (SVM) will have an important impact on the fault classification and performance of the support vector machine. Based on this, a fault analysis and diagnosis model using ant colony algorithm to optimize support vector machine is proposed to improve the accuracy of gearbox fault diagnosis. First, the collected original vibration signal is decomposed by EEMD to obtain the modal function component IMF, and then the energy entropy of the IMF component is calculated as the feature vector of the original vibration signal. Finally, the feature vector is input to the support vector optimized by the ant colony algorithm identify and classify in the machine, and finally get the diagnosis result. Comparing ACO-SVM with SVM, the experimental results prove that the ACO-SVM model has a higher fault diagnosis rate, stronger optimization ability, and faster convergence speed.

1. Introduction
Wind turbines are generally in harsh environments, resulting in a very high rate of failure of the gearbox. Once the gearbox fails, it will seriously affect the efficiency of power generation. Therefore, timely diagnosis and maintenance of gearbox faults are of great significance [1]. Because the fault signals of gearboxes are mostly nonlinear and non-stationary [2], ZH Wu and H proposed a new white noise auxiliary data analysis method-Ensemble Empirical Mode Decomposition (EEMD) [3], EEMD can effectively filter out the influence of white noise, while ensuring the continuity of each modal function in the time domain to reduce modal aliasing. SVM has its unique advantages in processing small sample data. It does not pursue the optimal solution obtained from infinite samples, but can find the optimal solution in limited samples. However, the kernel function g and penalty coefficient c in SVM are often artificially determined, and their value is extremely important for the recognition accuracy of SVM [4]. In recent years, swarm intelligence algorithm has made significant progress in optimizing SVM parameters. In this paper, ant colony algorithm is used to optimize SVM parameters [5]. Experimental results show that this method can obtain a higher fault recognition rate.

2. EEMD
2.1 The basic principle of EEMD
Using the random characteristic that the mean value of white noise is zero, a group of white noise is added to the original signal, and then the signal after the noise is decomposed by EMD, and the
corresponding overall average value of the obtained IMF is calculated as the result of EEMD decomposition. The addition of white noise ensures the continuity of the modal function in the time domain and reduces the phenomenon of modal aliasing\cite{6}. Specific steps are as follows:

(1) After adding m groups of random white noise sequences with equal zero variance to the original signal \(x(t)\), it becomes a group of m signals \(\{x_1(t), x_2(t), \ldots, x_m(t)\}\).

(2) Perform EMD decomposition of each signal \(x_i(t)(i = 1, 2, \ldots, m)\) in the signal group after adding noise to obtain m groups of IMF components \([c_{i1}, c_{i2}, \ldots, c_{in}],[c_{21}, c_{22}, \ldots, c_{2m}], \ldots, [c_{m1}, c_{m2}, \ldots, c_{mn}]\) (denoted as \(C\)) and m groups of residuals \([Re_{s_1}, Re_{s_2}, \ldots, Re_{s_m}]\) (denoted as \(Res\)).

\[
C = \left[ \begin{array}{c}
  c_{11} \\
  \vdots \\
  c_{1n} \\
  c_{21} \\
  \vdots \\
  c_{2n} \\
  \vdots \\
  c_{mn} \\
\end{array} \right],
Res = \left[ \begin{array}{c}
  Re_{s_1} \\
  \vdots \\
  Re_{s_1} \\
  Re_{s_2} \\
  \vdots \\
  Re_{s_2} \\
  \vdots \\
  Re_{s_m} \\
\end{array} \right]
\]

(3) Calculate the mean value of the m groups of IMF components \(\{c_1, c_2, \ldots, c_n\}\) and the mean value \(Res\) of the m groups of margin. Then the original signal can be expressed as: \(x(t) = \sum_{i=1}^{m} c_i + Re\). The realization process of EEMD is shown in Figure 1:

\[\text{Figure 1 Flow chart of EEMD method}\]

This article takes the pitting failure signal as an example, and introduces the signal waveform when a pitting failure occurs in the gearbox. Pitting failure signal is shown in Figure 2:

\[\text{Figure 2 Pitting failure signal}\]
The signal is decomposed by EEMD to obtain 13 IMF components and a margin Res. The result of EEMD decomposition is shown in Figure 3:

![EEMD decomposition result graph](image)

2.2 EEMD energy entropy

The gearbox obtains different vibration signals under different working conditions, and the energy distribution will also change with the different working conditions, so it can be judged whether the
gearbox is malfunctioning according to the different energy values under different conditions\cite{7}. In the process of EEMD decomposition to get the IMF, the frequency components in the original signal will also be decomposed into each IMF, each IMF component will have a different frequency of energy, the energy distribution corresponding to each IMF component is the energy entropy, which is assumed to be ignored. Res contains the influence of energy, then the sum of the energy of each IMF component is equal to the total energy of the original signal, so the EEMD energy entropy is defined as:

\[ H_{E_{xy}} = -\sum_{i=1}^{n} p_i \log^p \]

Among them \( p_i = \frac{E_i}{E} \) is the proportion of the energy of the i-th IMF in the total energy \( E = \sum_{i=1}^{n} E_i \).

It can be seen from the above analysis that the working status and fault type of the gearbox can be judged by the magnitude of the EEMD energy entropy value. The energy entropy value in the normal state is the largest. When the gearbox fails, the energy entropy value will decrease, so the energy can be reduced. Entropy is input into the ACO-SVM model as a feature vector to further judge the working state of the gearbox.

3. Ant Colony Algorithm
The ant colony algorithm mainly simulates the process of ants going out for food collectively in nature. At first, ants will start from the nest and foraging in all directions. After finding food, they will attract other ants to come. The ant will leave a pheromone on the path that has been traversed. The concentration of the pheromone will volatilize over time. Later ants will follow the pheromone on each path. The concentration of the path selection will lead to more ants walking on the path with the higher the pheromone concentration, forming a strong positive feedback\cite{8}.

Suppose that all m newborn ants are randomly placed in n new cities, and at the same time, The position of each ant in the taboo table is the position of the city where the ant is currently located. At this time, the pheromone concentration on each path is equal. Suppose \( \tau_k(0) = c \) (c is a small constant) that each ant will choose the next city to go to according to the pheromone concentration on the current path (the distance between the two cities \( p_{ij}(t) \)). For a city, the probability that ants k transfer from city i to city j at time t is:

\[ p_{ij}^k(t) = \frac{\left[ \tau_{ij}(t) \right]^\alpha \cdot \left[ \eta_{ij}(t) \right]^\beta}{\sum_{r \in J_k(i)} \left[ \tau_{ir}(t) \right]^\alpha \cdot \left[ \beta_{ij} \right]^\beta}, J \in J_k(i) \]

\[ 0, \text{ other} \]

Where: \( J_k(i) \) represents a set of city expectations that need to be re-selected after Ant K gets down, and the taboo table records the cities that Ant K has successfully walked through. \( \eta_{ij}(t) \) represents the expected value of ants transferring from a city i to a city j. \( \alpha \) And \( \beta \) denote the importance of pheromone and expectation heuristics, respectively. When all the ants have completed a tour, the pheromone and expected heuristic factors on the route of each city will be updated once according to the following formula:

\[ \tau_{ij}(t+n) = (1-\rho) \cdot \tau_{ij}(t) + \Delta \tau_{ij} \]

Where: \( \rho(0 < \rho < 1) \) represents the average evaporation time increment coefficient of each different pheromone time length on the path, \( 1-\rho \) represents the persistence length evaporation increment coefficient of each different pheromone, and \( \Delta \tau_{ij} \) represents each different information on the edge ij in this iterative transformation. The increase in the time evaporation length of the element.

4. Support vector machine optimized based on ant colony algorithm
Since the kernel function parameter g and penalty factor c of SVM are usually artificially set, they will
have an important impact on the function classification and performance of SVM. In view of the above problems, this paper proposes a gearbox fault diagnosis based on ant colony algorithm optimization support vector machine. The model uses the global optimization capability of ant colony algorithm to optimize SVM parameters to improve the fault diagnosis rate of SVM\cite{9}. The flowchart is shown in Figure 4:

![Figure 4 Flow chart of ant colony algorithm to optimize SVM](image)

5. Experimental results and analysis

5.1 Sample selection and feature extraction

In this paper, we use Jiangsu Qianpeng Diagnostic Engineering Co., Ltd. gearbox random failure analysis data, and randomly collect and obtain the random vibration failure signals of the gearbox in four states: normal vibration state, pitting vibration failure, broken tooth vibration failure, and wear vibration failure. In each state, 30 groups of vibration fault signals are randomly selected, 25 groups of which are randomly selected as fault training samples, and the remaining 5 groups are used as fault test samples. First, perform EEMD decomposition of the training sample, select the first 8 IMF signal components with abundant vibration fault information after the sample decomposition, and then calculate the energy entropy corresponding to each component to form a feature vector, and then input it into the support of the ant colony algorithm optimization For training in a vector machine, the energy feature vectors of the training samples are shown in Table 1 (due to space issues, only part of the feature vectors of each state are listed, and only 4 significant digits are retained). Finally, the test samples are used to compare the trained support vectors The machine performs the diagnosis of the working state of the gearbox.
Table 1 Feature vectors of some training samples

| Gearbox status   | Serial number | P1    | P2    | P3    | P4    | P5    | P6    | P7    | P8    |
|------------------|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| normal status    | 1             | 0.3546| 0.0389| 0.0423| 0.1476| 0.1367| 0.36045| 0.2526| 0.0227|
|                  | 2             | 0.3675| 0.0458| 0.0612| 0.1574| 0.1721| 0.3651 | 0.1631| 0.0181|
|                  | 3             | 0.3674| 0.3397| 0.1668| 0.0513| 0.0239| 0.0129 | 0.0065| 0.0034|
|                  | 4             | 0.3492| 0.3656| 0.1925| 0.0639| 0.0333| 0.0156 | 0.0145| 0.0203|
| Pitting failure  | 1             | 0.2046| 0.1174| 0.0758| 0.2318| 0.3412| 0.3486 | 0.1015| 0.0192|
|                  | 2             | 0.2191| 0.0882| 0.0758| 0.0835| 0.2675| 0.3104 | 0.3392| 0.0261|
|                  | 3             | 0.3363| 0.3652| 0.1217| 0.0346| 0.0215| 0.0082 | 0.0039| 0.0023|
|                  | 4             | 0.3299| 0.3669| 0.1852| 0.0577| 0.0318| 0.0119 | 0.0062| 0.0041|
| Broken tooth fault| 1           | 0.2434| 0.0271| 0.0356| 0.0918| 0.2032| 0.3036 | 0.3096| 0.0232|
|                  | 2             | 0.1081| 0.0169| 0.0655| 0.1187| 0.2063| 0.2922 | 0.3179| 0.0366|
|                  | 3             | 0.2014| 0.3077| 0.1228| 0.0364| 0.0167| 0.0071 | 0.0032| 0.0017|
|                  | 4             | 0.2396| 0.3259| 0.1503| 0.0473| 0.0261| 0.0108 | 0.0052| 0.0035|
| Wear failure     | 1             | 0.1484| 0.0182| 0.1086| 0.2576| 0.3255| 0.3639 | 0.2906| 0.0239|
|                  | 2             | 0.0958| 0.0465| 0.1082| 0.2587| 0.3104| 0.3679 | 0.3436| 0.0552|
|                  | 3             | 0.1447| 0.2393| 0.1212| 0.0383| 0.0172| 0.0084 | 0.0049| 0.0023|
|                  | 4             | 0.2815| 0.3549| 0.1640| 0.0495| 0.0261| 0.0107 | 0.0056| 0.0033|

5.2 Modeling and diagnosis

The ant colony algorithm is used to optimize the parameters of the support vector machine. The number of iterations of the ant colony algorithm is 300 and the number of population size is 50. When the number of iterations of the best fitness curve is the third generation, it becomes stable, indicating that it tends to converge and obtain the most Optimal selection of parameters. In order to verify the effect of the gearbox fault diagnosis model established in this paper, ACO-SVM and SVM are compared, and the results are shown in Table 2. It can be seen that ACO-SVM can search for the optimal value after a small number of iterations. Compared with traditional SVM, ACO-SVM has a faster convergence time and a higher fault diagnosis rate.

Table 2 Comparison results of ACO-SVM and SVM

| Data set        | Different algorithms | Operation hours | Convergence time | Accuracy  |
|-----------------|----------------------|-----------------|-----------------|-----------|
| Pitting failure | SVM                  | 155.4823        | 0.7654          | 96.40%    |
|                 | ACO-SVM              | 152.684         | 0.5643          | 98.85%    |
| Broken tooth fault| SVM                | 88.462          | 0.6754          | 96.50%    |
|                 | ACO-SVM              | 55.468          | 0.6653          | 96.80%    |
| Wear failure    | SVM                  | 78.456          | 0.2641          | 95.70%    |
|                 | ACO-SVM              | 77.894          | 0.1486          | 98.90%    |

It can be seen that in terms of running time, convergence time, and fault diagnosis accuracy, ACO-SVM has more advantages than SVM, which also proves the effectiveness and stability of the ACO-SVM model in gearbox fault diagnosis.

6. Conclusion

Aiming at the non-stationarity and nonlinearity of the vibration signal when the gearbox fails, and the influence of the support vector machine kernel function and penalty factor on the support vector machine, this paper uses EEMD to process the vibration signal when the gearbox fails, and uses the ant colony algorithm to The kernel function and penalty factor of the support vector machine are optimized for parameters, and an ACO-SVM model is established. The experimental results show that the ACO-SVM model has a faster convergence time and a higher fault diagnosis accuracy rate than the
SVM model. Therefore, the ACO-SVM model established in this article has extremely high application value and prospects.

References
[1] Lei, Y.G., He, Z.J., Lin, J., Han, D., Kong, D.T., (2011) Research progress of planetary gearbox fault diagnosis technology. J., Chinese Journal of Mechanical Engineering., 47(19): 59-67.
[2] Feng, Z.P., Chu, F.L., (2013) Torsional vibration signal analysis method for planetary gearbox fault diagnosis. J., Proceedings of the Chinese Society of Electrical Engineering., 33(14): 101-106.
[3] Ma, Z.Q., Li, Y.C., Liu, Z., Gu, C.J., (2016) Rolling bearing fault feature extraction based on variational modal decomposition and Teager energy operator. J., Vibration and Shock., 35(13): 134-139.
[4] Wang, H.Y., Li, J.H., Yang, F.L., (2014) Summary of support vector machine theory and algorithm research. J., Computer Application Research., 31(05): 1281-1286.
[5] Du, H.J., Li, Y., (2012) Research on the influence of parameter settings in ant colony algorithm on its performance. J., Modern Computer (Professional Edition)., (13): 3-7.
[6] Liu, H., Lin, J.H., Ding, J.M., Liu, X.C., (2017) Empirical Mode Decomposition Envelope Technology of AM-FM Signal and Modal Aliasing. J., Chinese Journal of Mechanical Engineering., 53(02):1-10.
[7] Wang, F.Z., Feng, W.C., Han, S.M., Yao, L., (2020) Motor bearing fault diagnosis based on EEMD energy entropy and hybrid algorithm. J., Measurement and Control Technology., 39(02): 44-49.
[8] Chen, L., Shen, J., Qin, L., (2003) A new approach to continuous parameter optimization with ant colony algorithm. J., System Engineering Theory and Practice., (03): 48-53.
[9] Gao, L.F., Zhang, X.L., Wang, F., (2015) Application of improved ant colony algorithm in SVM parameter optimization research. J., Computer Engineering and Applications., 51(13): 139-144.