A Fault Diagnosis Model of Marine Diesel Engine Fuel Oil Supply System Using PCA and Optimized SVM

Liangsheng Hou¹, Jundong Zhang¹ and Baisong Du²

¹Marine Engineering College, Dalian Maritime University, Dalian, Liaoning, 116026, China
²School of Port & Transportation Engineering, Zhejiang Ocean University, Zhoushan, Zhejiang, 316022, China

E-mail: zhjundong@dlmu.edu.cn

Abstract. The fuel oil supply system of the marine diesel engine contains many components, which fits plenty of sensors to monitor the condition of all components. A fault sample consists of data collected from all the sensors at certain time, which lead the dimension of the fault sample is very high. When the ship is sailing, there is a randomness in fault categories and fault duration, which leads the fault data unbalanced. This paper proposes an appropriate combinational approach to address the above problems. First, to reduce computational complexity, the high dimensional fault samples are converted into the low dimensional ones using the principal component analysis (PCA). Second, a sample size optimization (SSO) strategy is proposed to address the problem of the learning from the imbalanced datasets, which improve the classification performance of support vector machine (SVM). Third, a three-dimensional Arnold mapping is introduced into the particle swarm optimization (PSO) algorithm to improve its generalization capability. Finally, the SVM optimized by the improved PSO is trained as the classifier to identify the ten faults in the fuel oil supply system. Results demonstrate that the average correct diagnosis ratio can be as high as 93.9%.

1. Introduction

Shipping is the earliest form of transportation and plays a crucial role in total world transportation [1]. The power onboard is provided by the marine diesel engine, which is called the heart of the ship [2]. The fuel oil supply system of marine diesel engine provides adequate fuel oil to the engine, making sure the steady output power. Therefore, timely fault diagnosis of the fuel oil supply system is an important guarantee for the whole shipping.

Preventing accidents and ensuring the safety of the crew is the direct purpose and basic task of carrying out the fault diagnosis work. The first country to develop fault diagnosis technology is the United States. In 1967, NASA and the naval research institute established the first mechanical failure prevention team. Since then, fault diagnosis technology has developed rapidly. The development of fault diagnosis technology has gone through three stages: the original fault diagnosis stage, the stage of fault diagnosis based on sensor and computer technology and the intelligent fault diagnosis stage. In the original fault diagnosis stage, people relied on their experience to make direct judgments by the state of the device. This method was simple and practical in the fault diagnosis of simple equipment. However, people were required to have a wealth of knowledge and experience in a particular field. In the mid-1960s, fault diagnosis was performed by sensors and computer technology due to their
convenience in data measuring and processing. In the intelligent fault diagnosis stage, artificial intelligence technology was introduced to the fault diagnosis [3][4]. The intelligent fault diagnosis methods consist of case-based reasoning method [5][6][7], fault tree analysis method [8], rule-based reasoning method [9], fuzzy theory reasoning method [10], and neural network diagnosis method [11][12]. A growing number of successful cases regarded the intelligent fault diagnosis as a new direction in the development of current diagnostic technology [13][14].

In this study, we conduct the fault diagnosis on the marine diesel engine fuel oil supply system. The datasets collected from the marine diesel engine fuel oil supply system has two challenging problems, which is common in the industrial machines. One problem is the fault features are higher dimensional. Another is the datasets are unbalanced, which mean the fault samples in some classes are very scarcer than other classes. We propose a combinational method to address the above problems. First, we use PCA to converted the high dimensional fault samples into low-dimensional ones. Then, a sample size optimization strategy is proposed to help SVM to learn from the imbalanced datasets. Finally, the SVM optimized by the PSO which is modified by the chaotic mapping is trained as the classifier to perform the fault diagnosis of the fuel oil supply system.

The remainder of this paper is organized as follows. In Section 2, the fault diagnosis model is established. In Section 3, the fault diagnosis results of ship fuel oil system are analyzed and discussed. Finally, conclusions are given in Section 4.

2. Fault diagnosis model establishment

2.1. Fuel oil supply system of marine diesel engine

In the fuel oil system, the oil is stored in the tanks in the double bottom. After passing through the centrifuges, the oil is delivered to a service tank. Then the oil is transferred by a fuel oil supply pump to a fuel-water emulsion controller, from where the oil is pumped through a heater. The fuel oil circulation pump raises the pressure of fuel oil to a constant value before the main engine and diesel generators. The fuel oil is then passed through a fine filter before being supplied to the injection system. The fuel oil supplying system contains (1) Strainers, (2) Fuel oil supply pumps, (3) Fuel-water emulsion controller, (4) Venting box, (5) Fuel oil circulation pumps, (6) Fuel oil heaters, (7) Fuel oil back flush filter, and (8) Fuel oil bypass filter, as shown in Figure 1.

![Figure 1. Fuel oil supply system.](image-url)

2.2. Normal faults in the fuel oil supply system

In this part, we choose 31 common monitoring parameters in the fuel oil system. Namely, they are fuel oil circulation pump discharge pressure, fuel oil circulation pump suction pressure, fuel oil pressure outlet main engine to heater, fuel oil pressure at main engine, fuel oil filter differential pressure, fuel oil water content inlet main engine, fuel oil viscosity inlet main engine, fuel oil gassing indication, main engine pressure control valve position, fuel oil temperature inlet main engine fuel oil heater, fuel oil temperature outlet main engine fuel oil heater, fuel oil viscosity outlet main engine fuel oil heater,
steam pressure inlet control valve, main engine fuel oil heater outlet temperature, fuel oil supply pump pressure, fuel oil supply pump casting temperature, fuel oil venting box level, fuel oil venting box pressure, fuel oil venting box temperature, fuel oil venting box viscosity, fuel oil venting box gas content, fuel oil water content inlet main engine, fuel oil viscosity control valve position, fuel oil viscosity inlet main engine, fuel viscosity outlet main engine fuel oil heater, fuel oil venting box viscosity, fuel oil temperature outlet main engine fuel oil heater, main engine fuel oil heater steam pressure, main engine fuel oil heater temperature, main engine fuel oil heater outlet temperature, main engine fuel oil heater fuel oil pressure drop.

The common faults in the main engine fuel oil system include normal (F1), fuel oil circulation pump wear (F2), fuel oil circulation pump motor failure (F3), main fuel oil back flush filter dirty (F4), main engine fuel oil heater dirty (F5), fuel oil supply pump wear (F6), fuel oil supply pump motor failure (F7), fuel oil flow meter restriction (F8), fuel oil de-aerating valve malfunction (F9) and water in fuel oil to main engine (F10).

2.3. Data acquiring and preprocessing

In this paper, the fault samples are acquired from Kongsberg marine engine room simulator. We did 410 experiments and collected 410 fault samples. The collected fault data are imbalanced. 60% of the data are randomly chosen as training samples, and 40% are for testing. The ratio between the number of fault samples in F1, F3, F7 and F2, F4, F5, F6, F8, F9, F10 is 0.4.

In this paper, the raw data is preprocessed twice. Firstly, the raw sample matrix is normalized by Eq. (1).

\[
B_{ij} = \frac{A_{ij} - \mu_j}{\sigma_j}
\]

where \(A_{ij}\) are the elements of the \(i\)th row and the \(j\)th column in raw sample matrix \(A\), \(B_{ij}\) are the elements of the \(i\)th row and the \(j\)th column in normalized sample matrix \(B\), \(\mu_j\) is the expectation of \(j\)th column in raw sample matrix \(A\), \(\sigma_j\) is the standard variance of the \(j\)th column in raw sample matrix \(A\).

After the normalized sample matrix \(B\) are processed by PCA method, some principal components are selected as the new sample matrix \(C\), whose elements are processed by Formula (2) again.

\[
y = \frac{x - x_{\min}}{x_{\max} - x_{\min}}
\]

where \(x\) is the element in the sample matrix \(C\), and \(y\) is in the range of [0, 1], \(x_{\min}\) and \(x_{\max}\) are minimum and maximum values on each dimension of the fault data.

2.4. Optimize SVM using improved PSO

2.4.1. Initialize the population of PSO using chaotic systems. A random or predetermined strategy is used to initialize a population in standard PSO algorithm. However, the initial population may consist of inhomogeneous solutions. Based on the fact that PSO is extremely sensitive to the initial population, chaos variable is introduced to produce a uniformly distributed population. Arnold mapping has better chaotic properties comparing with other mappings. Here, a three-dimensional Arnold mapping is used to generate the initial population:

\[
\begin{align*}
x_{n+1} &= (x_n + y_n + z_n) \mod 1 \\
y_{n+1} &= (x_n + 2y_n + 2z_n) \mod 1 \\
z_{n+1} &= (x_n + 2y_n + 3z_n) \mod 1
\end{align*}
\]

where \(x_n\), \(y_n\) and \(z_n\) are the variables in the range (0-1).

The initial population can be obtained by the following procedure:

Step 1. Randomly generate the vectors \(x_1, y_1\) and \(z_1\) according to Eq.(4)-(6):

\[
x_1 = (x_{11}, x_{12}, \ldots, x_{1n}), \quad x_{1n} \in [0, 1]
\]
evaluate the \( \text{CDR} \) and fault misclassified ratio (FMR). At last, the running time of models is calculated to evaluate the complexity. All criteria are defined by Eq. (9) and (10).

\[
\text{CDR} = \frac{n_1}{n_{\text{total}}} \% \tag{9}
\]

\( y_1 = (y_{11}, y_{12}, \ldots, y_{1n}), \quad y_{1n} \in [0,1] \tag{5} \)

\( z_1 = (z_{11}, z_{12}, \ldots, z_{1n}), \quad z_{1n} \in [0,1] \tag{6} \)

Step 2. Generate \( M (M > m) \) chaotic variables: \( x_1, x_2, \ldots, x_M \) according to Eq. (3); Step 3. Calculate the fitness value of M chaotic variables, and choose the first m chaotic variables in the descending order of fitness values as the individuals of the population.

2.4.2 \textit{Optimize SVM using modified PSO}. The penalty factor \( c \) and kernel function parameter \( g \) should neither too small nor too large. The performance of the SVM model is most affected by \( c \) and \( g \). In this study, we use improved PSO to obtain the best value of \( c \) and \( g \). The correct diagnosis ratio is chosen as the fitness value of PSO. The initial population is generated by the three-dimensional Arnold maps.

2.5. \textit{Optimize SVM using sample size optimization strategy}
The performance of a trained SVM model is vital to the classification ratio of testing samples. Generally, the training samples with the same size can improve the performance of SVM. However, the sample size varies in each category. This could lead SVM to ignore some samples in a small sample size category when training SVM [15]. In this paper, we use SSO strategy to improve the performance of SVM.

Assuming that each category in the training samples has a different sample size. The repeating time \( n_i \) of all samples in the \( i \)th category can be calculated by Eq. (7). The total training samples of the \( i \)th category is calculated by Eq. (8) where \( \left( \frac{m_{\text{max}}}{m_i} \mod 1 \right) m_i \) samples are randomly chosen in the training samples in the \( i \)th category.

\[
n_i = \left\lfloor \frac{m_{\text{max}}}{m_i} \right\rfloor \tag{7}
\]

\[
m_i^{\text{train}} = \left\lfloor n_i m_i + \left( \frac{m_{\text{max}}}{m_i} \mod 1 \right) m_i \right\rfloor \tag{8}
\]

Where \( \lfloor \cdot \rfloor \) denotes the round function that rounds toward zero, \( \lfloor \cdot \rfloor \) refers to the round function that rounds to the nearest integer, \( \frac{m_{\text{max}}}{m_i} \mod 1 = \frac{m_{\text{max}}}{m_i} - \left\lfloor \frac{m_{\text{max}}}{m_i} \right\rfloor \), \( m_{\text{max}} \) is the samples size of the largest category size in training samples, \( m_i \) is the samples size of the \( i \)th category in training samples and \( m_i^{\text{train}} \) is the total training samples of the \( i \)th category. The flowchart of the combinational model based on PCA and optimized SVM can be pictured as shown in Figure. 2. First, the high dimensional fault samples are converted into the low dimensional ones using the PCA method. The initial penalty factor \( c \) and kernel function parameter gamma (\( g \)) of SVM are got from the population initiate by chaotic mapping. Then, the SVM is trained, and each particle fitness is calculated. If the condition is not met, the population will update the velocity and position of the particles until a new population with the best optimal \( c \) and \( g \) that meets the conditions is generated. The SVM with the optimal \( c \) and \( g \) are adopted for fault diagnosis.

3. Faults diagnosis and results analysis
In this study, all the algorithms are designed in the Matlab (R2018a) platform and carried out on a computer (Intel(R) Core(TM) i7-3770 CPU @ 3.40GHZ and 8.00 GB of RAM). All the simulations in this paper are repeated ten times.

3.1. \textit{Statistical performance}
To evaluate the performance of the proposed model, two criteria are considered: correct diagnosis ratio (CDR) and fault misclassified ratio (FMR). At last, the running time of models is calculated to evaluate the complexity. All criteria are defined by Eq. (9) and (10).
\[ FMR = \frac{n_2}{n_{total}} \% \]  \hspace{1cm} (10)

where \( n_1 \) is the total number of samples correctly diagnosed, \( n_2 \) denotes the number of samples misclassified into other classes, \( n_{total} \) is the total number of samples.

**Figure 2.** The flowchart of the fault diagnosis model based on PCA and optimized SVM.

### 3.2 The optimal number of principal components selection

In this paper, we use PCA method to reduce the dimensionality of the fault samples. The contribution rate is calculated for each subject on a given principal component (PC), and the sum contribution rate for a different number of principal components is obtained. Figure. 3 presents the sum contribution rate of principles. The sum contribution rate is over 85% when the PCs reach 10 and over 94% when the PCs reach 14. In this study, 14 PCs are chosen for the fault diagnosis.

**Figure 3.** The sum contribution rate for different number of principal components.

### 3.3 Fault diagnosis with proposed and compared methods

In this part, six algorithms are applied to ship fuel oil fault diagnosis. Figure. 4 shows the first test diagnosis results with optimized SVM. Figure. 4(a) presents diagnosis results using the optimized
SVM (O_SVM) method and Figure 4(b) displays results using the optimized SVM with SSO (O_SVM+SSO) method. The testing accuracy of the optimized SVM with SSO method is 89.63%, and it is higher than the optimized SVM method, which is 87.8%.

The first test diagnosis results with PCA and optimized SVM method are shown in Figure 5. Figure 5(a) presents diagnosis results using the PCA and optimized SVM (PCA+O_SVM) method. Figure 5(b) displays diagnosis results using the PAC and optimized SVM with SSO (PCA+O_SVM+SSO) method. The testing accuracy of the PCA and optimized SVM with SSO method is 94.51%, and it is higher than PCA and optimized SVM method, which is 91.46%.

Figure 4. Results of optimized SVM.

Figure 5. Results of PCA and optimized SVM.

Figure 6. Results of BP and RBF neural network.
Multi-layer perceptron (MLP) with BP algorithm added momentum factor and adaptive learning rate, and Radial Basis Function Network (RBFN) are chosen as comparisons. Table 1 presents the parameters of MLP and RBFN. The architectures of MLP are as follows: 31-15-10 three-layer. Figure. 6 presents the first test diagnosis results for MLP and RBFN. The testing accuracies of the MLP and RBFN method are 80.49% and 75.61%, respectively.

The average CDR, FMR, and run time of the six algorithms are summarized in Table 2. The average CDR of the proposed method is 93.9%, and it is higher than other methods, which are 77.93%, 80.55%, 88.05%, 89.63%, and 92.8%, respectively. The average FMR of the proposed method is 6.1%, and it is lower than other methods, which are 22.07%, 19.45%, 11.95%, 10.37%, and 7.2%, respectively. The average run time of the proposed method is 78.27s, and the others are 0.98s, 0.53s, 79.99s, 102s and 65.8s, respectively.

### Table 1. The parameters of MLP and RBFN.

| Parameter                  | Value |
|----------------------------|-------|
| MLP Learning rate          | 0.01  |
| Ratio to increase learning rate | 1.05  |
| Ratio to decrease learning rate | 0.7   |
| Momentum constant          | 0.9   |
| Iterations                 | 1000  |
| RBFN Spread of radial basis functions | 1     |
| Maximum number of neurons  | 25    |

### Table 2. Comparing of the statistical indexes of differential algorithms

| Method          | CDR   | FMR   | Run time (s) |
|-----------------|-------|-------|--------------|
| RBFN            | 77.93%| 22.07%| 0.98         |
| MLP             | 80.55%| 19.45%| 0.53         |
| O_SVM           | 88.05%| 11.95%| 79.99        |
| O_SVM +SSO      | 89.63%| 10.37%| 102          |
| PCA+O_SVM       | 92.8% | 7.2%  | 65.8         |
| PCA+O_SVM +SSO  | 93.9% | 6.1%  | 78.27        |

*O_SVM means SVM algorithms optimized by PSO algorithm which is optimized by Arnold mapping.

### 4. Conclusions

For the real ships, the fault samples are high dimensional and unbalanced in the fuel oil supply system of the marine diesel engine. The above problems will weaken the classification ability of traditional machine learning methods. This paper proposes an appropriate combinational approach to address the above problems. We use PCA to converted the high dimensional fault samples into the low dimensional ones. An SSO strategy is proposed to help the SVM to learn from the unbalanced fault data. Finally, we train an improved SVM optimized by a modified PSO to identify the ten faults in the fuel oil supply system of the marine diesel engine.

### Acknowledgements

This work is supported by Research on Intelligent Ship Testing and Verification ([2018]473).

### References

[1] Bicer Y and Dincer I 2018 *Int. J. Hydrogen Energy*. **43** 1179–1193

[2] Zamiatina N 2016 *Procedia Eng*. **134** 157–164

[3] Wu Y, Chen Z, Wu L, Lin P, Cheng S and Lu P 2017 *Energy Procedia*. **105** 1070–1076

[4] Praveenkumar T, Saimurugan M, Krishnakumar P and Ramachandran K 2014 *Procedia Eng*. **97** 2092–2098
[5] Olivier-Maget N, Negny S, Hétreux G and Le Lann J. 2009 Eur. Symp. Comput. Aided Process Eng., Elsevier pp. 345–350
[6] Yang B, Kwon Jeong S, Oh Y and Tan A 2004 Expert Syst. Appl. 27 301–311
[7] Zhao H, Liu J, Dong W, Sun X and Ji Y 2017 Neurocomputing. 249 266–276
[8] Mentes A and Helvacioglu I 2011 Ocean Eng. 38 285–294
[9] Dou D, Yang J, Liu J and Zhao Y 2012 Knowledge-Based Syst. 36 1–8
[10] Kang Y, Kim H, Heo G and Song S 2017 Expert Syst. Appl. 69 239–246
[11] Rezaniaiee Aqdam H, Ettefagh M and Hassannejad R 2018 Ocean Eng. 164 284–297
[12] Oliveira J, Pontes K, Sartori I and Embriruçu M 2017 Expert Syst. Appl. 84 200–219
[13] Kazemi S, Rahai A, Daneshmand F and Fooladi A 2011 Ocean Eng. 38 774–781
[14] Peimankar A, Weddell S, Jalal T and Lapthorn A 2017 Swarm Evol. Comput. 36 62–75
[15] Tian Y, Fu M and Wu F 2015 Neurocomputing. 151 296–303