Monitoring and Analysis of Vehicle Engine State Based on Data Mining Technology

Xiuling Wei, Chuanxiang Du

Xi'an Siyuan University; Xi'an 710038, China

Abstract. the state of vehicle engine is influenced by many factors. The changing characteristics are very complex. The current model can not get high accuracy vehicle engine condition monitoring results. In order to improve the monitoring precision of vehicle engine state, a vehicle engine state monitoring model based on data mining technology is proposed. Based on the current research status of vehicle engine condition monitoring are analyzed, pointed out the deficiencies of the current vehicle engine condition monitoring model, then using wavelet analysis to decompose the state of vehicle engine, and the chaotic processing for each decomposed component, finally data mining technology -- support vector machine to monitor the decomposition rate of chaos treated by data mining, and monitoring their results reconstructed. Compared with other models, the simulation results show that this model can describe the change characteristics of vehicle engine state from multiple angles and get more accurate monitoring results of vehicle engine condition, which has higher practical application value.

1. Introduction

With the increasing number of vehicles, vehicles become an important means of transport, because the vehicle engine is the main driving mechanism of the vehicle, the vehicle engine fault detection probability increases rapidly, so the vehicle engine condition monitoring has aroused widespread concern in the state, but the vehicle engine is affected by various factors, with the change of chaos, change is very complicated, how to improve the precision of vehicle engine condition monitoring is the focus in the research field of vehicle[1,2].

According to the vehicle engine condition monitoring problem, domestic and foreign scholars to conduct extensive and in-depth research, put forward many excellent vehicle engine condition monitoring model of [3], the current vehicle engine condition monitoring values can be divided into two parts: vehicle engine condition monitoring of the traditional theory of statistical test based on [4,5] model, linear regression model, which is mainly based on the theory of linear modeling if the vehicle engine, is a single linear variation, simple and easy to implement, but the actual vehicle engine condition change is not a simple linear change characteristics, also has the characteristics such as nonlinear, stochastic, time-varying, cause this kind of model of vehicle engine condition monitoring error [6,7]. The other is the vehicle engine condition monitoring model [8] based on the modern statistics theory. In which the neural network is widely applied, because the neural network has good nonlinear learning ability, it obtains an ideal vehicle engine condition monitoring result [9-11]. In the training process, neural network is constantly updated and updated, and the convergence speed is slow. It is easy to appear "over fitting" condition monitoring result of vehicle [12]. Support vector machine (SVM) is a new data mining technology. It only needs one iteration to complete the training of neural network, and the work has been greatly improved. Because the change of network flow value is very complex, the single model can not obtain high precision vehicle engine state monitoring results [13].

In order to improve the monitoring precision of vehicle engine state, a vehicle engine state monitoring model (WA-SVM) based on data mining technology is proposed. Firstly, using wavelet analysis (wavelet analysis, WA) of the vehicle engine condition multiscale decomposition, and

chaotic processing for each decomposed component, and then through the support vector machine to monitor the decomposition rate of chaos after the treatment, monitoring of their results are reconstructed, finally simulation experiments to compare with other models, in order to effectively and the superiority of WA-SVM verification.

2. Correlation Theory

2.1 Phase Space Reconstruction

Phase space reconstruction work idea: any component of the evolution of the system by interacting components, related components of the information hidden in the component evolution process, thus by analyzing the time sequence of a component, understand the dynamic characteristics of the original system, extract and recover the original system. The state time series of vehicle engine is: x(t),t=1,2,...,N, m is embedding dimension and τ is time delay, can reconstruct a multi-dimensional vector sequence of X(t), mining change law hidden in the state of the vehicle engine time series, the original power system restoration of vehicle engine condition.

$$X(t) = \{x(t), x(t+\tau), \cdots, x[t+(m-1)\tau]\} \tag{1}$$

In the formula, M=N-(m-1)τ , M is the phase point number.

2.2 Wavelet Analysis

The wavelet analysis can carry on the multi-scale refinement of the signal, including two parts of decomposition and reconstruction. The Mallat algorithm is a fast wavelet transform algorithm, which is specific:

$$\begin{cases} p_l^{(N)} = \sum_n h_0(\lambda - 2l) p_l^{(N+1)}, & N \geq 0, \ N \in Z \\ q_l^{(N)} = \sum_n h_1(\lambda - 2l) p_l^{(N+1)}, & N \geq 0, \ N \in Z \end{cases} \tag{2}$$

In the formula, $h_0(\lambda)$ and $h_1(\lambda)$ are filter coefficients; $p_l^{(N)}$ and $q_l^{(N)}$ are high frequency and low frequency filters; for decomposition scale.

The form of signal reconstruction is

$$p_\lambda^{(N+1)} = \sum_l h_0(\lambda - 2l) p_\lambda^{(N)} + \sum_l h_1(\lambda - 2l) q_\lambda^{(N)} \tag{3}$$

2.3 Support Vector Machine

To introduce the subject the authors begin by outlining SVM for function approximation. Let the given training data sets be represented by (x_i, y_i), i=1,2...,n, where $x_i \in R^d$ is an input vector, $y_i \in R$ is its corresponding desired output, and n is the number of training data. In SVM, the original input space is mapped into the high dimensional space called feature space, by nonlinear mapping $x \rightarrow g(x)$, Let $f(x)$ be the output of SVM corresponding to the input vector x, in the feature space, then a linear function is constructed:

$$f(x) = \omega^t g(x) + b \tag{4}$$

Where, ω is a coefficient vector and b is a threshold.

SVM learning can be obtained by the minimization of the empirical risk on the training data, and the ε-intensive loss function is used for the minimization of empirical risk. The loss function is defined as:

$$L^\varepsilon(x, y, f) = |y - f(x)|_\varepsilon = \max(0, |y - f(x) - \varepsilon|) \tag{5}$$

Where, ε is a positive parameter that allows approximation errors smaller than ε. The empirical risk is

$$R_{emp}(\omega) = \frac{1}{n} \sum_{i=1}^n L^\varepsilon(y - f(x_i)) \tag{6}$$

Other than the ε-intensive loss, SVM tries to reduce the model complexity by minimizing $|\omega|^2$. This can be described by slack variables ξ_i and $\hat{\xi}_i$, which measure training data x_i whose deviations exceed the constant ε. Subsequently, the SVM approximation is obtained as the following optimization problem.

$$\min \frac{1}{2}|\omega|^2 + C\sum_{i=1}^{n}(\xi_i + \hat{\xi}_i)$$

$$s.t.$$

$$y_i - f(x_i) \le \varepsilon + \xi_i, f(x_i) - y_i \le \varepsilon + \hat{\xi}_i, \xi_i, \hat{\xi}_i \ge 0$$

(7)

Where, C is a positive constant to be regulated. By using the Lagrange multiplier method, the minimization of formula (7) causes the problem of maximizing the following dual optimization

$$\max \sum_{i=1}^{n} y_i(\hat{\alpha}_i - \alpha_i) - \varepsilon \sum_{i=1}^{n} y_i(\hat{\alpha}_i - \alpha_i) -$$

$$\frac{1}{2}\sum_{i,j=1}^{n}(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j)K(x_i, x_j)$$

$$s.t.$$

$$\sum_{i=1}^{n}(\hat{\alpha}_i - \alpha_i) = 0, C \ge \hat{\alpha}_i, \alpha_i \ge 0$$

(8)

Where, $\hat{\alpha}_i$ and α_i are Lagrange multipliers, and kernel $K(x_i, x_j)$ is a symmetric function, which is equivalent to the dot product in the feature space, namely

$$K(x_i, x_j) = g(x_i)^t(x_j)$$

(9)

Here the Gaussian function is used as kernel

$$K(x, y) = \exp(-\frac{\|x - y\|^2}{2\sigma^2})$$

(10)

Some other kernel is are given below:

The polynomial kernel is $K(x, y) = (x \cdot y + 1)^d$ and the hyperbolic tangent kernel is $K(x, y) = \tan h(c_1(x \cdot y) + c_2)$.

By replacing $\beta_i = \hat{\alpha}_i - \alpha_i$ and relation $\hat{\alpha}_i \alpha_i = 0$, the optimization of formulae (7) is rewritten as

$$\max \sum_{i=1}^{n} y_i \beta_i - \varepsilon \sum_{i=1}^{n}|\beta_i| - \frac{1}{2}\sum_{i=1}^{n}\beta_i\beta_j k(x_i, x_j)$$

$$s.t.$$

$$\sum_{i=1}^{n}\beta_i = 0, -C \le \beta_i \le C$$

(11)

The learning results for training data set D can be derived from Esq. (11). Note that only some of the coefficients β_i are not zeros and the corresponding vectors z are called support vectors (SV). That is, vectors x_i whose corresponding coefficients $\hat{\alpha}_i - \alpha_i$ are not zero are SV. Then the approximation function is represented by Lagrange multipliers, namely

$$f(x) = \sum_{i=1}^{P}(\hat{\alpha}_i - \alpha_i)k(x_i, x_j) + b$$

(12)

It should be noted that P is the number of SV, and $f(x)$ is computed only from SV. Furthermore, the constant b can be determined as well.

The quality of SVM mode is strongly depends on the proper setting of parameters and the SVM approximation performance is sensitive to parameters. For Gaussian kernel, the parameters to be

regulated include hyper parameters C, ε, and kernel parameter σ. The problem of parameter selection is further complicated by the fact that the SVM model complexity (which is the generalization performance) depends on these three parameters. Indeed, the values of C, σ, and ε relate to the actual function model and they are not fixed for different data sets.

The values of parameter C, σ, and ε affect the model complexity in a different way. Parameter C determines the trade-off between model complexity and the degree to which deviations larger than ε are tolerated. Parameter ε controls the width of the ε-insensitive zone and can affect the number of SV in the optimization problem. Kernel parameter σ determines the kernel width and relates to the input range of the training data set.

3. State Monitoring Model of Vehicle Engine Based on WA-SVM

(1) the state of the vehicle engine is decomposed by wavelet, and the different components are obtained, and the phase space reconstruction of the different components is carried out.

(2) use the extreme learning machine to establish the corresponding vehicle engine state monitoring model for the reconstructed low frequency components and high-frequency components respectively, and get the monitoring results of low frequency components and high-frequency components.

(3) finally, the monitoring results of the components are reconstructed and fused, and the results of vehicle engine state monitoring are obtained.

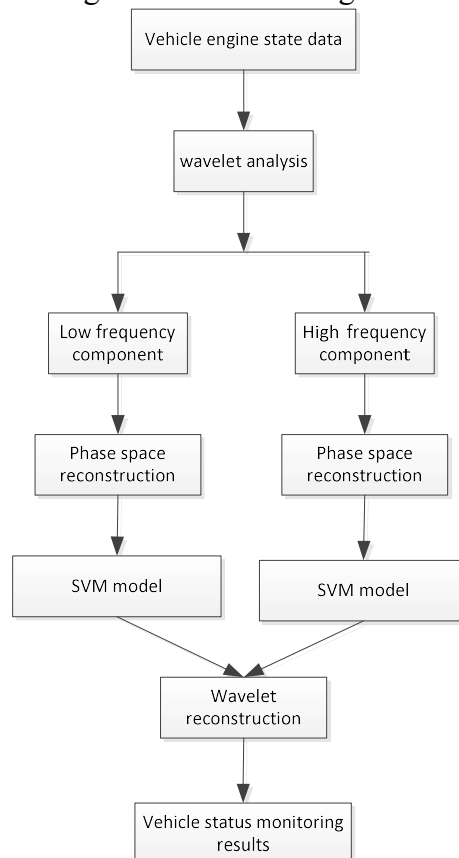The working flow of the vehicle engine state monitoring model of WA-SVM is shown in Figure 1

Fig.1 WA-SVM vehicle engine state monitoring flow

4. Simulation Experiment

4.1 Data Source

In order to test the performance of the vehicle engine state monitoring of WA-SVM, 1000 samples are collected for each hour of the vehicle engine state as an experiment, as shown in Figure 2. 100 samples were selected as test samples and other samples were used as training samples.
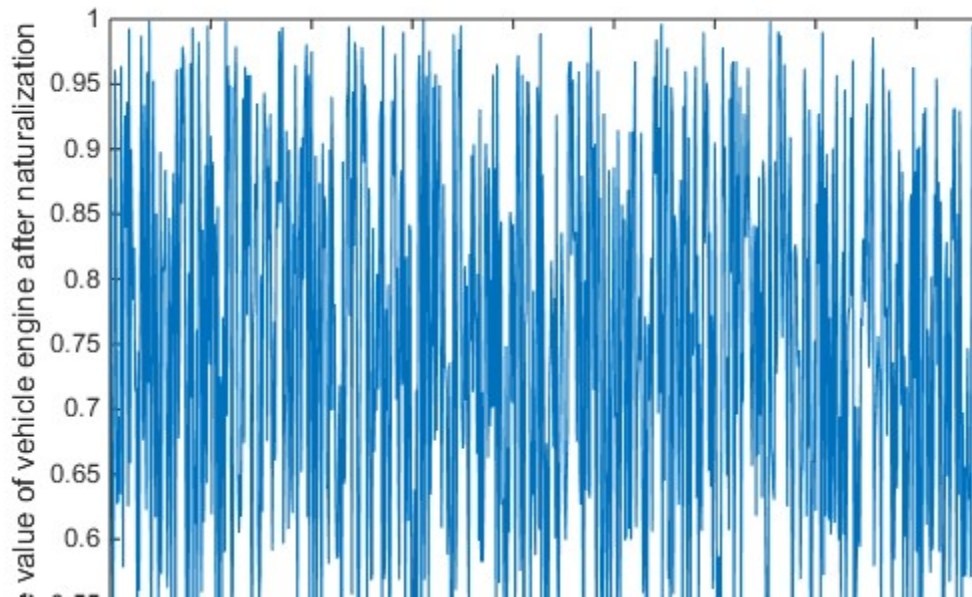
Fig.2 experimental data

Single support vector machine (SVM) and wavelet analysis +BP neural network (WA-BPNN) were selected as comparison models. The average relative percentage error (MPAE) and root mean square error (RMSE) were used to evaluate the monitoring results.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}(i)-y(i)}{y(i)}\right|\times 100\% \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}(i)-y(i))^2} \tag{14}$$

In the formula, $y(i)$ and $\hat{y}(i)$ are the expected value and the actual monitoring value; N is the number of test samples.

4.2 Construction of Learning Samples

Mallat algorithm using wavelet analysis to decompose the vehicle engine condition, low frequency and high frequency sequence a1, d1, d2, and then estimate the vehicle engine condition tau and m, as shown in table 1. According to table 1 τ and m respectively, Reconstruction the state of the vehicle engine time series of the feature, the various components of the learning sample.

Table 1 the components of M and tau value

Component name	m	τ
a1	2	7
d1	4	7
d2	3	8

4.3 Results and Analysis

The SVM components of the vehicle engine condition for learning, establish the monitoring model of each component, each component of the monitoring results obtained by wavelet reconstruction, the fusion of the monitoring results to get the final results of the vehicle monitoring component, engine condition, single step fitting and the monitoring results as shown in figure 2. From Figure 2, we can see that WA-SVM can effectively fit the change characteristics of vehicle engine state, and accurately depict the trend of vehicle engine state. It is a vehicle engine state monitoring model with high accuracy and fitting and monitoring.
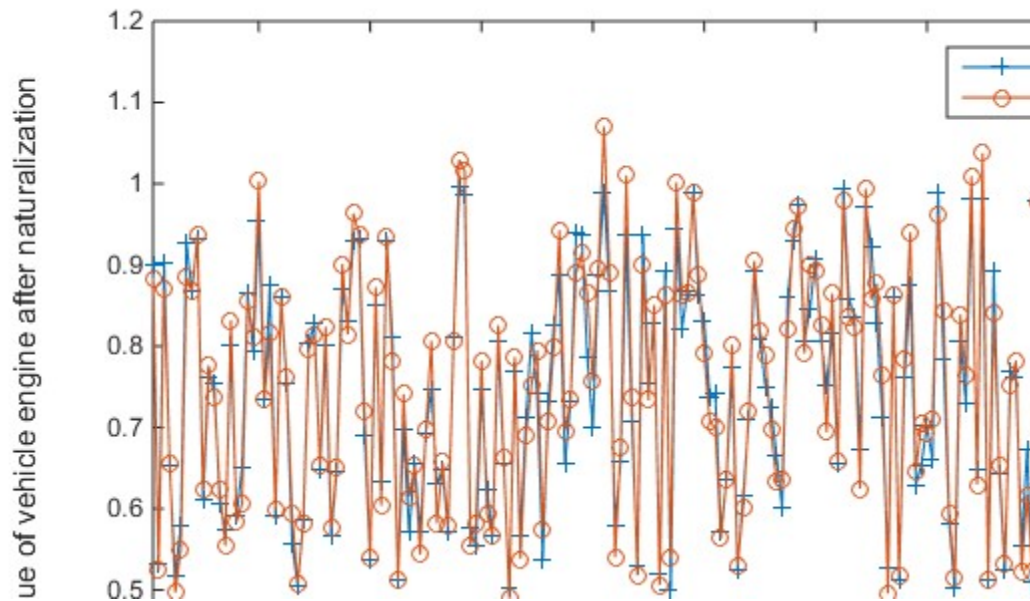
Fig.2 WA-SVM monitoring results

The monitoring errors of SVM, WA-BPNN and WA-SVM are shown in Table 2. From table 2, we can see that the MAPE and RMSE of WA-SVM are all less than SVM and WA-BPNN. This is because the wavelet analysis can subdivide the vehicle engine state, and help the modeling of vehicle engine state. SVM can better reflect the trend of vehicle engine state.

Tab.2 monitoring performance comparison

model	RMSE	MAPE(%)
SVM	24.09	10.25
WA-BPN	21.34	9.23
WA-SVM	19.24	7.55

5. Conclusion

The vehicle engine condition is the combined result of various factors, so it is stochastic, periodic and chaotic characteristics of vehicle engine condition monitoring results can help managers understand the change trend of vehicle engine condition in advance, for it is important to study. In view of the fact that the current model can not accurately describe the state change of vehicle engine, a WA-SVM vehicle engine condition monitoring model is proposed, and the following conclusions are obtained through simulation experiments.

(1) use wavelet analysis to decompose the vehicle engine state data, get more subtle characteristics of vehicle engine state change, help to excavate the change state of vehicle engine state, and improve the monitoring accuracy of vehicle engine condition.

(2) use chaos theory to reconstruct phase space of vehicle engine state decomposition, and change the time relationship between vehicle state data based on one-dimensional time-varying data and multi-dimensional time series, which is conducive to subsequent vehicle engine state modeling and monitoring.

(3) the support vector machine is applied to learn the state data of vehicle engine, and an ideal vehicle engine condition monitoring model is established, which has better vehicle engine condition monitoring results than other models, and has higher practical application value.

Acknowledgements

References

[1]. akthivel N, Sugumaran V, Nair B B. Application of Support Vector Machine (SVM) and Proximal Support Vector Machine (PSVM) for fault classification of mono-block centrifugal pump [J]. International Journal of Data Analysis Techniques and Strategies, 2010, 2(1): 38-61.

[2]. Salahshoor K, Kordestani M, Khoshro M S. Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers [J]. Energy, 2010, 35(12): 5472-5482.

[3]. Wenyi L, Zhenfeng W, Jiguang H, et al. Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM [J]. Renewable Energy, 2013, 50:1-6.

[4]. Ni J, Zhang C, Yang S X. An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs [J]. Power Delivery, IEEE Transactions on, 2011, 26(3): 1960-1971.

[5]. Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines [J]. Machine Learning, 2002, 46(1-3): 131-159.

[6]. Xiao J-Z, Wang H-R, Yang X-C, et al. Multiple faults diagnosis in motion system based on SVM [J]. International Journal of Machine Learning and Cybernetics, 2012, 3(1): 77-82.

[7]. Chen Bo, Liu Xinjian, Yuan love people. The application of [J]. Internal combustion engine and power device, fault tree analysis method in fault diagnosis of marine main engine 2010, 6): 48-50.

[8]. Gao Xin, Liu Xiangxuan, Cheng Kun. Failure fault tree analysis of diesel engine lubricating oil [J]. internal combustion engine, 2013, 2): 58-60.

[9]. Wang Mingjie, Yang Ling. Armored vehicle autoloader fault tree fuzzy analysis theory [J]. automation based on 2010,29 (5): 81-82.

[10]. Zheng Bo. Research on SVM fault diagnosis based on Gauss kernel function [J]. Journal of China Civil Aviation Institute of flight, 2012, 5): 49-52.

[11]. Jiang Wan, Wu Shengqiang. Multi data fusion fault diagnosis method based on SVM and evidence theory [J]. Instrument and instrument journal, 2010, 8): 1738-1743.

[12]. Lv Jianxin, Wu tiger Sheng, Wu Lushan, et al. based on EMD complexity characteristics and SVM bearing fault diagnosis research [J]. Mechanical transmission, 2011, 35 (2): 20-23.

[13]. Shi Zhibiao, Song Quangang, Ma Mingzhao, et al. based on improved particle swarm optimization support vector machine for steam turbine fault diagnosis [J]. Dynamic Engineering Journal, 2012, 32 (6): 454-457.