

Prediction Method of Equipment Degradation State Based on Improved RVM

Lu Cheng^{1,a}, Wang RuiQi¹, Xu TingXue¹ and Chen YuQi¹

¹ Coastal Defense College, NAAU Yantai, China

Abstract. In order to improve the prediction accuracy of the relevance vector machine model, an improved method for equipment condition prediction is proposed. First of all, an improved kernel function of variance Gauss kernel (VGKF) is constructed to improve the global performance and generalization ability of the kernel function. Then, by using the method of selecting the number of adjacent points in the chaotic sequence local prediction method, the H-Q criterion was used to optimize the embedding dimension of the training space to avoid the blindness of subjective selection. Through the prediction example of terminal guidance radar equipment test parameters, the effectiveness and superiority of the improved RVM were verified.

1 Introduction

State prediction technology is the key technology of condition-based maintenance strategy. It integrates the current and historical status monitoring data of equipment to determine its status and development trend in the future, so as to provide decision-making information for maintenance support[1].

In recent years, the research of equipment condition prediction mainly focuses on grey theory, artificial neural network, cloud theory and correlation vector machine[2-5], but the equipment history test data usually has the characteristics of small sample, nonlinear and dynamic uncertainty, and the traditional theory and method have some limitations.

Relevance vector machine[6-7] is a new machine learning algorithm based on sparse Bayes learning theory proposed by TIPPING in 2001. Compared with SVM, its kernel function is not limited by Mercer conditions, so it is more flexible to select; the parameter setting is simple, and the complex additional parameters are avoided; the sparsity of the solution is much higher than that of SVM, and it has better generalization ability.

Based on the above analysis, in order to further improve the prediction accuracy of the correlation vector machine model, an improved RVM equipment condition prediction method is proposed. Finally, the prediction performance of the improved RVM is verified by an example of terminal guidance radar equipment.

2 RVM Prediction Model

^aCorresponding author: kvcelu@163.com

RVM is a nonlinear sparse Bayesian learning theory. Its good generalization performance and less correlation vector make it a better application in prediction field. The principle of the regression model is as follows:

Let be a training dataset with the sample number of N , $\mathbf{x}_i \in \mathbb{R}^D$ is D dimension input vector, $t_i \in \mathbb{R}$ is output scalar. Assuming that they are all independently distributed, the relationship between the two is represented as follows:

$$t_i = y(\mathbf{x}_i, \omega) + \varepsilon_i \tag{1}$$

Among them, $\varepsilon_i \sim N(0, \sigma^2)$ is sample Gaussian noise. An expression similar to SVM, which can be represented by a series of kernel functions:

$$y(\mathbf{x}, \omega) = \sum_{i=1}^N \omega_i K(\mathbf{x}, \mathbf{x}_i) + \omega_0 \tag{2}$$

Among them, $\omega = [\omega_0, \omega_1, \dots, \omega_N]^T$ is the weight parameter vector, and $K(\bullet, \bullet)$ is the kernel function. So it can be deduced that $p(t_i | \mathbf{x}_i) = N(t_i | y(\mathbf{x}_i, \omega), \sigma^2)$, t_i satisfies $N(y(\mathbf{x}_i, \omega), \sigma^2)$. In order to facilitate the expression and introduce a super parameter, the likelihood function of the entire training sample data set is represented as:

$$p(\mathbf{t} | \omega, \beta) = \left(\frac{\beta}{2}\right)^{N/2} \exp\left\{-\frac{\beta}{2} \|\mathbf{t} - \Phi\omega\|^2\right\} \tag{3}$$

Among them, $\mathbf{t} = [t_1, t_2, \dots, t_N]^T$, $\Phi \in \mathbb{R}^{N \times (N+1)}$ is the design matrix, defined as $\Phi = [f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_N)]^T$, the base function vector is $f(\mathbf{x}_i) = [1, K(\mathbf{x}_i, \mathbf{x}_1), \dots, K(\mathbf{x}_i, \mathbf{x}_N)]^T$, $i = 1, 2, \dots, N$.

The training of sample data using RVM is to obtain the posterior distribution of the weight parameter vector ω . It is assumed that the prior distribution of ω_j satisfies the Gaussian distribution with a mean value of 0 and a variance of \dots , expressed as:

$$p(\omega | \alpha) = \prod_{j=0}^N N(\omega_j | 0, \alpha_j^{-1}) \tag{4}$$

Among them, each α_j in the hyper parametric $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]$ is independent of each other and is only related to the corresponding weight ω_j . By means of formula (3) and formula (4), the mean matrix μ and the covariance matrix \hat{a} , respectively:

$$\hat{a} = (\beta\Phi^T\Phi + A)^{-1} \tag{5}$$

$$\mu = \beta\hat{a}\Phi^T\mathbf{t} \tag{6}$$

among them, $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$. To obtain the posterior distribution of ω , two impact parameters β and α_j must be optimized, the specific method is to maximize the edge likelihood function $p(\mathbf{t} | \alpha, \beta)$. After obtaining the objective function on both sides of the $p(\mathbf{t} | \alpha, \beta)$ equal sign, the partial derivative of α_j and β is obtained according to the objective function, and the derivative is 0, and the iterative formula of two parameters is obtained:

$$\alpha_j = \frac{1}{\mu_j^2 + \sum_{ij} \gamma_j} = \frac{\gamma_j}{\mu_j^2}, j = 0, 1, \dots, N \tag{7}$$

$$\beta = \frac{N - \sum_{j=0}^N \gamma_j}{\|\mathbf{t} - \Phi\mu\|^2} \tag{8}$$

$$\gamma_j = 1 - \alpha_j \sum_{ij} \tag{9}$$

Among them, μ_j is the j element of $\boldsymbol{\mu}$, and \sum_{jj} is the j diagonal elements of covariance matrix $\hat{\mathbf{a}}$.

The training of the RVM model is to update the $\boldsymbol{\mu}$ and $\hat{\mathbf{a}}$ by iterative computation until the parameter converges until the parameter converges or reaches the maximum number of iterations. In the convergence of the parameter estimation, $\boldsymbol{\alpha}_{MP}$ and σ_{MP} are obtained by the maximum likelihood method. Given the new input value, \mathbf{x}_* , the predicted mean y_* and variance σ_*^2 are:

$$y_* = \boldsymbol{\mu}^T \phi(\mathbf{x}_*) \tag{10}$$

$$\sigma_*^2 = \sigma_{MP}^2 + \phi(\mathbf{x}_*)^T \hat{\mathbf{a}} \phi(\mathbf{x}_*) \tag{11}$$

3 Parameter Optimization of RVM Regression Model

3.1 Improvement of Gauss kernel function

Under the prior structure, the RVM algorithm removes the irrelevant sample points by using the auto correlation judgment theory, thus realizing the sparseness of the model [8]. Although compared with SVM, the kernel function of RVM is independent of the restriction of Mercer condition, but the performance of the predefined kernel function will directly affect the accuracy of RVM regression prediction.

The commonly used kernel functions include Gauss kernel function, Laplace kernel function, Cauchy kernel function and so on. Gauss kernel function is the most widely used kernel function, and

its standard form is defined as $K_{std}(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2b^2}\right\}, i = 1, 2, \dots, N$, among them, the parameter b is the nuclear width, and N is the training sample number.

As a kernel function based on the distance equation, the test points of Gauss kernel function tend to become sparse when they are mapped from the source space to the high-dimensional feature space. The above can be changed only when the nuclear function is not only able to maintain a rapid attenuation at the test point, but also to maintain a certain attenuation at infinity. However, the Gauss kernel function only satisfies the former, so it is a typical local kernel function. In recent years, many scholars composed of mixed kernel by constructing Gauss kernel function and global kernel function to improve the performance of kernel function [9-11], but in this process, multiple parameter variables are introduced, which greatly increases the computational complexity of RVM. Based on the Gauss kernel function, this paper improves the performance of RVM by constructing an improved variance Gauss kernel function.

It can be observed that in the standard form of Gauss kernel function, the characteristics of all elements of the scale factor uses the same measure. This feature also makes the characteristics of large numerical scale range more dominant, and this limitation seriously affects its performance in practical applications. To overcome this problem, we normalize all the feature elements by zero mean, and mark them as unit standard deviations:

$$\bar{x}_i = \frac{x_i - \mu_i}{\sigma_i} \tag{12}$$

Among them, μ_i and σ_i are the mean and standard deviations of the eigenvector, and the formula (12) is substituted into the standard form, and the result is rewritten into the form of the kernel function, so a new kernel function is generated:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left\{-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma_i^2 \bullet b^2}\right\}, i = 1, 2, \dots, N \tag{13}$$

Since σ_i^2 is the variance of the vector i , we call formula (13) the variance Gauss kernel function, so that the eigenvectors are characterized by different scaling factors. Further analysis makes it possible for the function to maintain a slow decay rate while away from the sample points, consider putting the Euclid distance equation on the denominator of the exponential function so that the improved VGKF expression follows:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp \left\{ -\frac{2\sigma_i^2 \bullet b^2}{\|\mathbf{x} - \mathbf{x}_i\|^2} \right\}, i = 1, 2, \dots, N \tag{14}$$

The kernel function constructed in this way not only meets the two conditions mentioned above, but also avoids the introduction of unknown parameters. Moreover, the normalization process of data is not needed before the input of model data.

3.2 Optimization of training sample space dimension

The key to realize the time series prediction of $\{x_1, x_2, \dots, x_n\}$ is to find the accurate mapping of the sliding time window between $X_t = \{x_{t-1}, x_{t-2}, \dots, x_{t-m}\}$ and output $Y_t = \{X_t\}$, the mapping is $f: R^m \rightarrow R$, $t = m+1, m+2, \dots, n$, m is used to predict embedding dimension. Therefore, the determination of parameter m is crucial for predicting the accuracy of modeling. In order to overcome the blindness of subjective m selection, the local chaotic time series prediction of the number of neighboring points selection method based on the idea of training sample space[12], the optimal embedding dimension of RVM were optimized, the method is suitable for small amount of data, has simple calculation and high precision.

According to the idea of small sample training in local method[13-14], the predictive value of target points can be inferred by the motion trend of adjacent phase points, that is to approximate the evolution value of current vector $X(n)$ and future value \hat{x}_{n+T} by mapping, the approximate process of this mapping is achieved by relying on the d adjacent point $X(n)$ of $X(n_i)(i = 1, 2, \dots, d)$. In this paper, we consider the use of neighborhood points to train RVM, and propose a method for optimizing the spatial dimension of RVM training samples. $X(n_i)(i = 1, 2, \dots, d)$ is used as input sample, and $x_{n_i+1}(i = 1, 2, \dots, d)$ is used as output value to train RVM regression prediction, then the weight coefficients and correlation vectors are obtained. In this way, the local linear model is replaced by the RVM regression model, and the good approximation performance of RVM is fully utilized. After training, an optimal function is obtained to fit $f(\bullet)$, and the prediction value of the subsequent data can be obtained by calculating \hat{x}_{n+T} .

In the process of determining the number of adjacent points, this paper is based on the Hannan-Quinn criterion[15], and the process is as follows:

First, a broad range of d is given to the $d \in [d_{\min}, d_{\max}]$, and the HQ criterion values for each d value are calculated sequentially:

$$C(d) = \ln \sigma(d)^2 + (d+1) \frac{H \ln(\ln N)}{N} \tag{15}$$

$$\sigma^2(d) = \frac{\frac{1}{S} \sum_{j=1}^S (x_j - x'_j)^2}{\bar{x}^2} \tag{16}$$

Among them, H is the constant representing the weight, $H > 2$ is generally used, N is the number of samples to be predicted, σ^2 is the fitting variance, x_j is the sample data point, \bar{x} is the sample mean, and x'_j is the sample prediction value, When the minimum value of formula (15) is reached, the

corresponding d_{opt} value is the optimum number of adjacent points. At the same time, the model achieves the best balance of accuracy and complexity in theory.

In this process, the spatial dimension of training samples is optimized by explicit $m = d_{opt}$, which overcomes the shortcomings of previous subjective training based on subjective experience and FPE criterion, and has certain reference and practical value.

4 Forecasting Examples and Analysis

Take the missile in the storage state of the troops as an example, from the beginning of 2009 to the end of 2016, the terminal guidance radar unit test was carried out every quarter, and 32 sets of test data of 2 key parameters were obtained, as shown in Table 1. Based on this, the validity of the improved RVM model was verified.

Table 1. Test Data of Key Parameter

Number	Prediction parameter	
	c_1	c_2
1	4.47	28.35
2	4.46	28.69
...
31	4.13	34.37
32	4.02	34.95

The previous 25 sets of data were used as training data for RVM, and the latter 7 sets of data were used to test the predictive effect of the model. After the embedding dimension is determined by the H-Q criterion, the RVM is trained by 21 training samples after phase space reconstruction. Using the improved RVM and the standard RVM model (the Gauss kernel width is set to 4), 2 performance parameters are predicted, as shown in Figure 1 and figure 2.

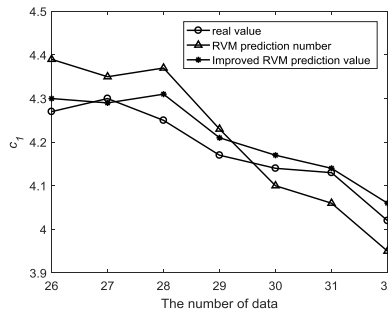


Figure 1. Prediction curves of parameter c_1

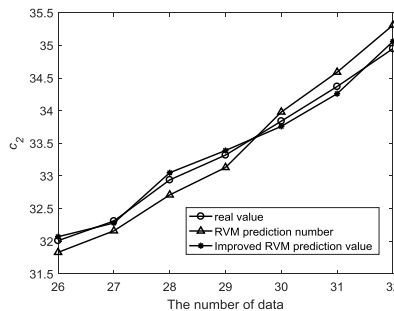


Figure 2. Prediction curves of parameter c_2

Furthermore, the average prediction error of the two is compared, and the average absolute percentage error and the correlation vector are used as the predictive performance evaluation indicators, as shown in table 2.

Table 2. Comparison of Prediction Results

Prediction method	c_1		c_2	
	MAPE/%	nRV	MAPE/%	nRV
RVM	1.81	6	0.63	5
Improved RVM	0.75	4	0.24	4

As can be seen from table 2, the improved RVM's prediction MAPE is lower than the standard RVM model, and the number of correlation vectors is also reduced. It shows that it has better prediction performance, which verifies the effectiveness and superiority of the proposed method.

5 Conclusion

In order to improve the prediction accuracy of equipment status, the RVM model is improved in this paper. By constructing the variance Gauss kernel function, the global performance and generalization ability of the kernel function are improved. Then, the H-Q algorithm is used to optimize the embedding dimension of the training space to avoid the blindness of its subjective selection. The performance prediction example of the terminal guidance radar proves that the improved RVM has higher prediction accuracy, and it is easy to calculate and has a good prospect of engineering application. At present, the prediction accuracy of RVM model is limited by single kernel function, In the future, the RVM based on the combination kernel function will be the focus of the next research.

References

1. Ma Sa Sa, Jia Xisheng, Xia Lianghua, "Review of military equipment maintenance engineering CBM," Journal of Institute of equipment, vol.19, pp. 111-116, February 2008.
2. Yingfeng L, "The Application of Improved Grey Model in Power System Load Forecast," Power System and Clean Energy, vol.3, pp. 10-11, 2009.
3. DongQin L, "Application of Support Vector Machine Regression Method in Prediction of Ship's Price Index," Ship & Ocean Engineering, vol.38, pp.104-106, February 2009.
4. Zhu Jianping, "Research on Application of neural network in power load forecasting," Science and technology information, vol.13, pp. 32, August 2015.
5. Zhao Wenbang, Wang Shaohua, Pu Ying, etc, "Prediction of equipment technical status based on cloud theory," Fire Control & Command Control, vol.38, pp.62-65, July 2013.
6. Tipping M E, "Sparse Bayesian learning and the relevance vector machine," Journal of machine learning research, vol.1, pp.211-244, June 2001.
7. Samui P, Dixon B, "Application of support vector machine and relevance vector machine to determine evaporative losses in reservoirs," Hydrological Processes, vol.26, pp.1361-1369, September 2012.
8. Zhou Jianbao, Research on prediction method of remaining life of lithium ion battery based on RVM, Harbin: Harbin Institute of Technology, 2013.
9. Yu J, "A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes," Chemical Engineering Science, vol.68, pp.506-519, January 2012.
10. Wei L, Wei B, Wang B, "Text Classification Using Support Vector Machine with Mixture of Kernel," Journal of Software Engineering and Applications, vol.5, pp.55, December 2013.

11. Duan Qing, Zhao Jianguo, Ma Yan, "Optimized combination kernel function correlation vector machine model for power load forecasting," *Journal of electrical engineering and control*, vol.14, pp.33-38, June 2010.
12. Du Jie, Cao family, Liu Zhijian, etc, "Multi step prediction model of local high-order Volterra filter for chaotic time series," *Acta phys Sinica*, vol.58, pp.5997-6005, September 2009.
13. Qing-Fang M, Yue-Hui C, Yu-Hua P, "Small-time scale network traffic prediction based on a local support vector machine regression model," *Chinese Physics B*, vol.18, pp2194, June 2009.
14. Bing Q, Gong B, Yang Z, et al. "Short-Term Traffic Flow Local Prediction Based on Combined Kernel Function Relevance Vector Machine Model," *Math. Probl. Eng*, vol.9, pp.1-9, October, 2015.
15. Hannan E J, Quinn B G, "The determination of the order of an autoregression," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.190-195, 1979.