Research on bearing life prediction based on support vector machine and its application

Chuang Sun, Zhousuo Zhang*, Zhengjia He
State Key Laboratory for Manufacturing Systems Engineering, Xi’an Jiaotong University, 710049, Xi’an, P.R. China
E-mail: sun.chuang@stu.xjtu.edu.cn   zzs@mail.xjtu.edu.cn
* Corresponding author.

Abstract. Life prediction of rolling element bearing is the urgent demand in engineering practice, and the effective life prediction technique is beneficial to predictive maintenance. Support vector machine (SVM) is a novel machine learning method based on statistical learning theory, and is of advantage in prediction. This paper develops SVM-based model for bearing life prediction. The inputs of the model are features of bearing vibration signal and the output is the bearing running time-bearing failure time ratio. The model is built base on a few failed bearing data, and it can fuse information of the predicted bearing. So it is of advantage to bearing life prediction in practice. The model is applied to life prediction of a bearing, and the result shows the proposed model is of high precision.

1. Introduction
Rolling element bearings are among the mechanical parts used widely, and are easily damageable [1-2]. The running condition of bearing plays a crucial role in the performance of mechanical equipment. Bearing failure can lead to costly loss in production and even human casualties. Effective prediction of bearing life is necessary to prevent abrupt bearing breakdown, and is also benefit to improve productivity, reduce costs and repairing time.

Accurately estimating the life of bearing is a challenging problem. In the conventional life prediction methods, the statistical model is predominant. Statistical model is based on an array of bearing fatigue life tests, and can reflect the randomness of rolling contact fatigue and the probability distributions of bearings life. It is significant for a batch of bearings life prediction. However, due to the randomness of bearing failure time, it is lack of accuracy for life prediction of single bearing in practice.

In recent years, modeling methods of bearing life prediction based on the condition monitoring data, such as temperature, vibration and sound emission signal, have obtained more and more attention. These models are data-driven models, and they can establish the non-linear relationship between condition monitoring data and actual bearing operating time directly. Of all the bearing monitoring data, vibration signal is more effective and suitable for reflecting bearing running condition [3]. The vibration is small and smooth when the bearing is under normal condition. And the occurrence of bearing defect can cause fluctuation of vibration. During the process of degradation, the amplitude of vibration increases obviously. Thus vibration signal becomes the convenient variable to investigate the degeneration process of bearing.
For accurate prediction of the bearing failure time, we develop a data-driven model based on SVM. In the model, principal component analysis (PCA) is used to select the vibration signal features of rolling element bearing run-to-failure test. Particle swarm optimization (PSO) is applied to the parameters optimization of SVM. After the processes of feature selection and parameter optimization, life prediction models are constructed based on support vector machine regression. The inputs of the models are features of vibration signal and the output is a scalar related to bearing life. The model can reflect the relationship between multiple features of vibration signal and bearing life better. The information of the predicted bearing is fused in the model, so the model is more accurate.

The remainder of this paper is organized as follows. Section 2 presents an introduction of support vector machine regression. Section 3 presents bearing life prediction models briefly. Section 4 describes the process of experiment validation. The conclusion is shown in section 5.

2. Support Vector Machine Regression

Support vector machine (SVM) based on statistical learning theory aims at limited samples classification and regression problems [4]. It was put forward by Vapnik during the 1990s. SVM has excellent generalization ability by utilizing the structural risk minimization principle to replace empirical risk minimization principle. Summary about the SVM regression is given in this section.

Given the training sample sets \((x_i, y_i)\), for \(i = 1, 2, \cdots, m\), where \(m\) is the number of the samples, \(x_i \in \mathbb{R}^n\) is a \(n\)-dimension vector, \(y_i \in \mathbb{R}\) is a scalar. In case of linear regression, the regression function takes the following form:

\[
f(x) = \langle \omega, x \rangle + b
\]

where \(\omega \in \mathbb{R}^n\) is a coefficient vector and \(b \in \mathbb{R}\) is a bias. The vector \(\omega\) and bias \(b\) are used to define the regression function \(f\). The optimal regression function can be acquired as a solution to the following optimization problem:

\[
\begin{aligned}
\text{minimize} & \quad \frac{1}{2} \|\omega\|^2 \\
\text{subject to} & \quad y_i - \langle \omega, x_i \rangle - b \leq \varepsilon \\
& \quad \langle \omega, x_i \rangle + b - y_i \leq \varepsilon
\end{aligned}
\]

where \(\varepsilon\) denotes prediction precision. The assumption in (3) is that the prediction error is less than \(\varepsilon\) for all the training samples. Considering the situation that not all the samples can meet formula (3), we introduce slack variables \(\xi_i, \xi_i^*\) and error penalty \(C\) to cope with the optimization problem. Hence the optimizing target and constraint conditions can be expressed as

\[
\begin{aligned}
\text{minimize} & \quad \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \\
\text{subject to} & \quad y_i - \langle \omega, x_i \rangle - b \leq \varepsilon + \xi_i \\
& \quad \langle \omega, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\
& \quad \xi_i, \xi_i^* \geq 0
\end{aligned}
\]

constant \(C\) determines the trade-off between the flatness of \(f\) and the amount up to which deviations larger than \(\varepsilon\) are tolerated [5]. By application of Lagrangian multiplier and saddle point condition this problem can be converted into dual optimization problem as follows:

\[
\begin{aligned}
\text{maximize} & \quad \left\{ -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*) \right\} \\
\text{subject to} & \quad \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0 \text{ and } 0 \leq \alpha_i \leq C, 0 \leq \alpha_i^* \leq C
\end{aligned}
\]
and $b$ can be obtained by solving the dual optimization problem. And $\boldsymbol{\omega}$ can be completely described as a linear combination of the training samples. Hence, the regression function can be expressed as follows:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\langle x_i, x \rangle + b$$  \hspace{1cm} (8)$$

For nonlinear regression tasks, SVM can also be used by means of kernel function. By using the non-linear vector function $\phi(\cdot)$, the original data can be mapped onto high dimensional feature space, where the linear regression is possible. The linear regression function in dual form is expressed by

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\langle \phi(x_i), \phi(x) \rangle + b$$  \hspace{1cm} (9)$$

By applying a kernel function $K(x, x)$, the regression function in the feature space without explicit evaluation of $\phi(\cdot)$ will become

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b$$  \hspace{1cm} (10)$$

Any function that satisfies Mercer’s theorem [6] can be used as a kernel function. Different kernel functions, such as linear, polynomial, Gaussian RBF and wavelet kernel function, may be used in SVM. In this research, Gaussian RBF function is utilized, and the Gaussian RBF function can be expressed as following:

$$K(x, x) = \exp\left(-\frac{\|x - x\|^2}{(2\sigma^2)}\right)$$  \hspace{1cm} (11)$$

where $\sigma$ is the parameter of Gaussian RBF function.

3. Bearing life prediction model
In this section, the life prediction model is presented. There are two main steps included in the model. The first is construction of the bearing degradation model and the second is process of prediction using the model.

3.1. Construction of degradation model
In this section SVM is used to build degradation assessment model of bearings by means of vibration signal. The main steps for modeling are shown in Figure 1. There are three main steps in the model, namely, the selection of sensitive features, constructing inputs and output of SVM prediction model and the optimization of SVM parameters. The sensitive feature selection method proposed in [7] is utilized here to gain sensitive features that are obviously changing in the period of bearing degradation. The inputs and output of SVM are constructed by combination of sensitive features and the ratio of bearing running time to bearing failure time. The SVM parameters are obtained using the particle swarm optimization algorithm which is introduced in detail by [8]. The main processes of the degradation model can be described as following.

Considering a bearing that is undergoing run-to-failure test, we can store the vibration signal and the corresponding bearing running time together during its degradation. After a set of bearings run-to-failure tests, degradation information and failure time of each failed bearing can be recorded in a database. Assume that $N$ bearings degradation information is included in the database, using the degradation data of $j$th bearing $B_j$, $j=1,2,\cdots,N$, a SVM life prediction model can be trained, called as SVM$_j$. During the training of SVM$_j$, we can select $l$ sensitive features by means of PCA. A $l$-elements vector $F_j = [f_{j1}^n, f_{j2}^n, f_{j3}^n, \cdots, f_{jl}^n]$, which represents the $l$ amplitudes of the sensitive features for bearing $B_j$ at the $nth$ sampling point, is served as the inputs of SVM$_j$. The corresponding output is a scalar that is proportion of operating time at the $nth$ sampling epoch $t_j^n$ to $B_j$’s life $T_j$. A
parameter $p$ is introduced to represent the proportion, namely, $p_{ij}^n = t_{ij}^n/T_j$. The SVM prediction model can be constructed utilizing the combination of sensitive features and the proportion, namely, $(F_{ij}^n; p_{ij}^n)$. In different sampling point, we can obtain different combination of $F_{ij}^n$ and $p_{ij}^n$. During the degradation of the bearing, a sequence including $(F_{ij}^n; p_{ij}^n)$ can be constituted. The train set and test set of SVM can be selected from the sequence respectively. Furthermore, particle swarm optimization algorithm is used for selecting optimal parameters of SVM. The optimization process is going in the form of iteration, and when the testing error is smaller than the predetermined threshold, the training process is over. According to the $N$ bearing degradation data, $N$ SVM prediction models, $SVM_1$, $SVM_2$, …, $SVM_N$, can be built to assess the degradation process of the corresponding bearing.

![Flowchart](image)

**Figure 1.** Flowchart of constructing SVM prediction model.

### 3.2. Prediction processes using the degradation model

Considering a running bearing $B_j$ that is partially damaged, we can predict the failure time of the bearing at any time during its degradation. The main steps for prediction are shown in Figure 2. For example, if we want to predict the bearing failure time at the $kth$ sampling epoch, the sensitive features vectors can be calculated from the $lth$ sampling epoch to the $kth$ sampling epoch, namely, $(F_1^l, F_2^l, \cdots, F_k^l)$. Then we input the $k$ amplitude vectors $F_1^l, F_2^l, \cdots, F_k^l$ to all the SVM, $j = 1, 2, \cdots, N$. The output $p_{ij}^1, p_{ij}^2, \cdots, p_{ij}^k$ can be regarded as the corresponding proportion of operating time to the bearing life of $B_j$ at $lth$ to $kth$ sampling epochs. The operating time $t_1^l, t_2^l, \cdots, t_k^l$ are known, and the bearing life can be acquired through dividing the operating time by the output, namely, $t_1^l/p_{ij}^1, t_2^l/p_{ij}^2, \cdots, t_k^l/p_{ij}^k$. The average of $k$ predicted bearing life is treated as failure time of $B_j$ predicted by SVM, at the time of $kth$ sampling epoch; call it $T_j^k$.

The next step is to calculate weights of SVM, using the testing error value. The testing error associated with SVM, is represented as $e_j$. The weight can be defined as
where $j = 1, 2, \ldots, N$. The weights sum to unity, and SVM with small testing error is endowed with large weight. The final failure time predicted at the time of $k$th sampling epoch is calculated according to a weighted average of $N$ failure time predicted by each SVM $j$:

$$T^k = \sum_{j=1}^{N} w_j T^k_j$$

Figure 2. Flowchart of SVM models used for prediction.

4. Experiment Validation

In this section, we use bearing run-to-failure test data to validate the effectiveness of the proposed bearing life prediction method. The data that is analyzed here was downloaded from Prognostics Center of Excellence (PCoE) shared by Center for Intelligent Maintenance System (IMS), University of Cincinnati [9].

During the test, four bearings are installed on one shaft, as presented in figure 3. The rotation speed of the bearings is kept constantly at 2000 rpm. Constant radial load of 6000 lbs is put onto the shaft by a spring mechanism. Rexnord ZA-2115 double row bearings that have 16 rollers in each row are installed on one shaft. The vibration signal is recorded in the same time interval with the sampling rate 20 kHz and data length 20,480 points.

Figure 3. Bearing test rig and sensor placement sketch.

(a) Sensor placement sketch. (b) Test rig structure sketch.
In three groups of tests, bearing 3 and bearing 4 fail with an inner race defect and a roller element defect respectively in test one. Bearing 1 fails with an outer race defect in test three. The training bearing set consist of bearing 4 in test one and bearing 1 in test three are used to build SVM models. The life of bearing 3 that is used as a test bearing in test one is predicted using the two models.

Sensitive features are selected firstly. In this paper, 16 original feature parameters in time-domain, as shown in table 1, are extracted from vibration signals.

Table 1. Formula of time-domain feature parameters.

| Formula | Description |
|---------|-------------|
| $F_1 = \frac{1}{N} \sum_{i=1}^{N} x(i)$ | $F_1$ |
| $F_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (x(i))^2$ | $F_2$ |
| $F_3 = \frac{1}{N} \sum_{i=1}^{N} \sqrt{x(i)}$ | $F_3$ |
| $F_4 = \frac{1}{N} \sum_{i=1}^{N} x(i)$ | $F_4$ |
| $F_5 = \frac{1}{N} \sum_{i=1}^{N} (x(i) - F_1)^2$ | $F_5$ |
| $F_6 = \frac{1}{N} \sum_{i=1}^{N} (x(i) - F_2)^2$ | $F_6$ |
| $F_7 = \frac{1}{N} \sum_{i=1}^{N} (x(i) - F_3)^2$ | $F_7$ |
| $F_8 = \frac{1}{N} \sum_{i=1}^{N} (x(i) - F_4)^2$ | $F_8$ |
| $F_9 = \max(x(i))$ | $F_9$ |
| $F_{10} = F_k - F_9$ | $F_{10}$ |
| $F_{11} = \frac{F_2}{F_4}$ | $F_{11}$ |
| $F_{12} = \frac{F_3}{F_4}$ | $F_{12}$ |
| $F_{13} = \frac{F_1}{F_4}$ | $F_{13}$ |
| $F_{14} = \frac{F_1}{F_3}$ | $F_{14}$ |
| $F_{15} = \frac{F_2}{F_3}$ | $F_{15}$ |
| $F_{16} = \frac{F_6}{F_2}$ | $F_{16}$ |

where $x(i)$ is a signal series, for $i = 1, 2, \ldots, N$, $N$ is the number of data points.

In order to reduce feature dimension and improve the performance of prediction, sensitive features are chosen by PCA method. According to the proposed selection scheme in [7], $F_3, F_7, F_8$ are selected as sensitive features. Then the vector $F$ and $(F; p)$ can be constructed, where $F$ includes three elements corresponding to the three amplitudes of the sensitive features and $p$ is the corresponding ratio of bearing running time to bearing life.

PSO algorithm is utilized to optimize the parameters of SVM. The strategy with constriction weight is utilized and constriction weight $\chi$ is set to 0.7298. $c_1$ and $c_2$, which are often called as acceleration constants, are set to 2 equally. The maximum speed of particle is limited in 20% of the size of solution space. The ranges of SVM parameters variation are set as $C \in (0, 1000], \varepsilon \in [0.0001, 1]$ and $\sigma \in (0, 10]$. According to the above processes, two bearing life prediction models, namely, SVM$_1$ and SVM$_2$, can be built. The optimal parameters for SVM$_1$ and SVM$_2$ is shown in table 2.

Table 2. Parameters of SVM$_1$ and SVM$_2$.

|  | SVM$_1$ | SVM$_2$ |
|---|--------|--------|
| $C$ | 21.8223 | 496.8226 |
| $\varepsilon$ | 0.0563 | 0.0002 |
| $\sigma$ | 1.1451 | 1.4057 |

The test results of the two models are shown in figure 4. The x-coordinate are serial number of the test sample and y-coordinate is the ratio of bearing running time to bearing failure time. The average test errors of SVM$_1$ and SVM$_2$ are 0.093 and 0.067 respectively. Figure 4 shows that the changing trends of the actual value and the predicted value are similar, and most of the errors are very small except a few points. Prediction results indicate that the two models reflect the actual situation of corresponding bearing degradation reasonably.
The next step is life prediction for the test bearing based on the two models. 10 sampling points are chosen from the defect period of test bearing. The bearing life at the 10 points is predicted by SVM\(_1\) and SVM\(_2\) respectively, and the results are listed in table 3.

**Table 3.** Predicted life by two models at ten sample points (h).

|   | 1   | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10  |
|---|-----|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| SVM\(_1\) | 791.6 | 791.9 | 794.2 | 797.1 | 799.8 | 801.5 | 802.8 | 803.6 | 804.1 | 804.9 |
| SVM\(_2\) | 790.1 | 789.9 | 789.9 | 790.2 | 790.6 | 791.2 | 791.6 | 791.8 | 791.9 | 792.5 |

The weights of the two models can be calculated according to formula (12). The weight of SVM\(_1\) is 0.420 and SVM\(_2\) with a weight equivalent to 0.580. Then the weighted bearing life can be achieved according to formula (13), and is listed in table 4.

**Table 4.** Weighted bearing life at ten points (h).

|   | 1   | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10  |
|---|-----|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| bearing life | 790.7 | 790.8 | 791.7 | 793.1 | 794.4 | 795.5 | 796.3 | 796.7 | 797.1 | 797.7 |

The bearing can be treated as failed when the root mean square of vibration signal increases abruptly at the later defect period. The bearing failure time of the test bearing is 820.0 hour. To validate the accuracy of the predicted life, the errors are calculated between the predicted failure time and actual failure time according to formula (14):

\[
\text{error} = \left| t_r - t_p \right| / t_r \times 100\%
\]

where \(t_r\) is the actual failure time, \(t_r = 820.0\) and \(t_p\) is the predicted failure time.

The errors of the predicted failure time are shown in figure 5. The x-coordinate is the running time of the test bearing at the ten predicted points and y-coordinate is the error of predicted failure time. Figure 5 presents that all the errors are less than 4%, and the average error is 3.2%.
The bearing life calculated by $L_{10}$ formula is

$$L_{10} = \frac{10^{6}}{60 \times 2000} \times \left( \frac{23400}{6000} \right)^{\frac{10}{7}} = 778.1 \text{ hour}$$  \hspace{1cm} (15)

the error of $L_{10}$ life is 5.1%. The $L_{10}$ formula is a general bearing life prediction method available in common working condition, and it can not fuse the degradation information of bearing working under specified conditions. The bearing life prediction model proposed in this paper can fuse actual running information of bearing, so the prediction result is superior to $L_{10}$ formula.

5. Conclusion
To predict the bearing life accurately, the prediction model based on SVM is presented. The model is based on actual bearing condition monitoring data and it can reflect bearing degradation process reasonably. The average life prediction error of the test bearing is 3.2%, while the error of $L_{10}$ life is 5.1%. The prediction result shows that the proposed model is more accurate than $L_{10}$ formula.

Acknowledgement
This paper is supported by the Natural Science Foundation of China “Research on Hybrid Intelligent Technique and Its Application in Fault Diagnosis Based on Granular Computing” (Grant No.50875197).
Many thanks to Centre for Intelligent Maintenance System (IMS) in University of Cincinnati for sharing the bearing test data.

References
[1] Yuna P Jin Ch and Lei G 2009 Robust bearing performance degradation assessment method based on improved wavelet packet-support vector data description Mechanical Systems and Signal Processing 23 669–81
[2] Nagi G Mark L R Liu and Vijay P 2004 Residual Life Predictions From Vibration-Based Degradation Signals: A Neural Network Approach IEEE Transactions on Industrial Electronics 51 694-70
[3] Runqing H Lifeng X and Xinglin L 2007 Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods Mechanical Systems and Signal Processing 21 193–207
[4] V.N. Vapnik 1995 *The Nature of Statistical Learning Theory* (New York: Springer)

[5] Alex J S and Bernhard S 2004 A tutorial on support vector regression *Statistics and computing* **14** 199–222

[6] Cristianini N and Shawe N J 2000 *An Introduction to Support Vector Machines* (Cambridge: Cambridge University Press)

[7] Arnaz M and Robert X G 2004 PCA-Based Feature Selection Scheme for Machine Defect Classification *IEEE Transactions on Instrumentation and Measurement* **53** 1517-25

[8] Kennedy J and Eberhart R 1995 Particle swarm optimization *Proceedings of the 4th IEEE International Conference on Neural Networks* **1** 1942-8

[9] Jie L and Hai Q 2007 Rexnord Technical Services, ‘Bearing Data Set’, IMS, University of Cincinnati, NASA Ames Prognostics Data Repository, <http://ti.arc.nasa.gov/project/prognostics-data-repository>, NASA Ames, Moffett Field>, CA.