

Residual Life Prediction for Rolling Element Bearings Based on an Effective Degradation Indicator

Shuai Zhang · Yongxiang Zhang · Danchen Zhu

Submitted: 29 April 2015 / in revised form: 26 July 2015 / Published online: 20 August 2015
© ASM International 2015

Abstract Effective degradation indicator and robust prediction model are very important for residual life prediction. Thus a new residual life prediction based on Markov indicator and support vector is proposed. Since the Markov model is good at dealing with stochastic characteristics in time domain, Markov model is joined with multiple fault features for the construction of an effective degradation indicator of rolling element bearings. The support vector regression is used to construct an adaptive prediction model composed of two prediction models that are, respectively, based on historical data and online data. Thus the ultimate prediction result is obtained by taking a weighted average of the two prediction results captured by the two prediction models, and the weights are adjusted by the LMS to enhance the prediction accuracy. The experimental results show that the Markov indicator is more sensitive than the common features, and the proposed prediction method is more effective in comparison to other methods.

Keywords Rolling element bearings · Degradation indicator · Markov model · Support vector regression · Adaptive prediction · LMS

Introduction

Rolling element bearings that suffer from the harsh working condition are the most important components of the

rotating machine. Thus a precision residual life prediction is required, so that the fatal damage can be prevented and the personal safety can be guaranteed. Therefore, it deserves much to do some research on the improvement of the residual life prediction accuracy for rolling element bearings.

Residual life prediction is the important part of the PHM (prognostics and health management). And two key important elements that are degradation indicator and prediction model consist in residual life prediction. A sensitive degradation indicator can be helpful to timely detection of the incipient fault and the determination of the incipient threshold, while a reasonable prediction model can improve the prediction accuracy. Therefore, the key to improve the prediction accuracy is to select the sensitive degradation indicator and build the prediction model reasonably. In general, some fault features, such as kurtosis, RMS, peak-to-peak value, etc., are always used to evaluate the degradation performance. But these features are only effective for certain defect at certain stage. Furthermore, for the highly stochastic degradation these features can neither be sensitive to the incipient faults nor make good trend of the degradation performance. To address these problems, many methods have been developed. Qiu developed a new degradation indicator based on self-organizing map neural network, while Jihong Yan used a BP neural network-based indicator to evaluate the degradation performance [1, 2]. Yu developed some new indicators based on generative topographic mapping (GTM), Gaussian mixture model (GMM), and Bayes [3, 4]. Liao used genetic programming to improve the degradation performance [5]. The above methods can show some sensitive to incipient fault than the general fault features, but they do not consider the stochastic characteristics which can

S. Zhang · Y. Zhang (✉) · D. Zhu
Naval University of Engineering Power Engineering Marine
Engineering, Wuhan 430033, China
e-mail: zyx6302@aliyun.com

exhibit a great effect on the degradation performance in time domain. In addition to the development of the degradation indicator, lots of prediction models have also been proposed for residual life prediction. In [6–8], the authors proposed a new prediction method based on neural network which improved the long-term prediction accuracy slightly. However, the neural network has some limitations: (1) difficulty of determining the network structure and the number of nodes and (2) slow convergence of the training process. In [9, 10], the authors proposed some prognostic approach based on Bayesian theory, which can predict the probability distribution of the residual life. Meanwhile, in [11] the authors also proposed a new method that can obtain the residual life by computing Phase-Type (PH) distribution. But the shortcoming of this method is also very obvious. It needs a large number of accurate data of prior probability distribution. Unfortunately, it is difficult to meet in actual application. Marcos E. proposed a fault prediction method based on particle filtering [12]. In [13, 14], the authors also used particle filtering to predict residual life and obtained good results. However, the previous prediction models that seldom compromised the information of the real-time data were mostly built by historical data. But the real-time data can directly reflect the current development trend of the bearings’ residual life, while the historical data contain the empirical information of the bearings’ residual life. To this point, using both historical data and real-time data can exhibit some advantages theoretically in terms of efficiency and accuracy.

CHMM (Continuous Hidden Markov Model) that has been widely used in many fields can deal with stochastic characteristics very well and exhibit great effectiveness in comparison to the above-mentioned methods. In [15, 16], the authors used CHMM for fault diagnostic. Moreover, SVR (Support Vector Regression) as a prediction model has also been used in RUL prediction. In [17, 18], SVR was used for fault prognostic.

Therefore, an adaptive method based on Markov indicator and support vector is proposed. The CHMM is used for the construction of the degradation indicator for its ability of dealing with stochastic characteristics in time domain. SVR that can handle the small sample well is used to build the prediction model which is composed of two models, respectively, based on historical data and online data. At last, the ultimate prediction result is calculated by taking a weighted average of the two prediction results captured by these two prediction models, and the weights are adjusted by the LMS (Least Mean Square algorithm) to enhance the prediction accuracy. The predicted results indicate that the proposed method is more effective in comparison to other common prediction methods.

The Principle of the Adaptive Prediction Method

The prediction method is first used to find an effective indicator for degradation performance. Then the prediction model is used to predict the trend of the indicator. At last, the predicted result can be obtained according to the pre-determined threshold. In this paper, we use CHMM and SVR to build, respectively, the degradation indicator and the prediction model.

The Principle of the Degradation Indicator Based on CHMM

A sensitive indicator with significant trend is needed for monitoring. Then the failure threshold can be determined so that a severity fault can be prevented at incipient stage. For the low effectiveness of the general fault features, CHMM is used for the construction of the indicator so that the stochastic problem in time domain can be solved.

Hidden Markov Model (HMM) is a powerful statistical tool and has been applied in many areas. The essential construction of HMM is given as follows. We denote that S is the state alphabet set and O is the observation alphabet set:

$$S = (s_1, s_2, \dots, s_N), \quad O = (o_1, o_2, \dots, o_M), \quad (\text{Eq 1})$$

where N and M , respectively, denote the total number of the states and the observations. Then, the transition matrix A is given as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix}.$$

Thus

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i), \quad 1 \leq i, j \leq N, \quad (\text{Eq 2})$$

where q_t denotes the state at time t , while q_{t+1} denotes the state at time $t + 1$, and the formula denotes the probability that the state s_i at time t transmits to s_j at time $t + 1$. Then we denote the observation matrix B as follows:

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1M} \\ b_{21} & b_{22} & \dots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \dots & b_{NM} \end{bmatrix}. \quad (\text{Eq 3})$$

Thus $b_{ik} = P(\theta_t = o_k | q_t = s_i), 1 \leq i \leq N, 1 \leq k \leq M$, which denotes the existence probability of the o_k in state S_i . π is the initial probability array given as follows:

$$\pi = (\pi_1, \pi_2, \dots, \pi_N)$$

and

$$\pi_i = P(q_0 = s_i), \quad 1 \leq i \leq N. \quad (\text{Eq 4})$$

As the summary of the above discussion, the HMM can be given by $\lambda = (\pi, A, B)$. Then CHMM is given by the HMM with continuous observation matrix, $B = \{b_{jo}\}$, which are denoted by Gaussian mixture model as follows:

$$\begin{aligned} b_{jo} &= \sum_{k=1}^M c_{jk} N(o, \mu_{jk}, U_{jk}) \\ &= \sum_{k=1}^M \frac{c_{jk}}{(2\pi)^{D/2} |U_{jk}|^{1/2}} \exp \left[-\frac{(o - \mu_{jk})(o - \mu_{jk})^T}{2U_{jk}} \right]. \end{aligned} \quad (\text{Eq 5})$$

$j = 1, 2, \dots, N$

Of which, the definitions of the variables are as follows: M : the number of the Gaussian function; c_{jm} : the weights of the m th Gaussian function in state S_j , which satisfy $c_{jm} \geq 0$ $\sum_{m=1}^M c_{jm} = 1$, $j = 1, 2, \dots, N$; N : Gaussian probability density function; o : the observation sequences with the size $D \times T$, where D denotes dimension and T denotes the length of the sequences; μ_{jm} : the mean value vector of the m th Gaussian probability density function in state S_j ; and U_{jm} : the mean value vector of the m th Gaussian probability density function in state S_j .

Hence, CHMM can be denoted by $\lambda = (\pi, A, C, \mu, U)$.

For the purpose of constructing the degradation indicator, we introduce the evaluation problem of HMM. Given a HMM and a sequence of observations, the likelihood probability of the observation sequence can be computed. This problem could be viewed if a given observation sequence is generated by the given model. Thus we can use the above probability to solve the problem, which is given by

$$\begin{aligned} P(O|\lambda) &= \sum_Q P(O|Q, \lambda) P(Q|\lambda) \\ &= \sum_{q_1 \dots q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \dots a_{q_{T-1} q_T} b_{q_T}(o_T). \end{aligned} \quad (\text{Eq 6})$$

The computation of the above equation can be solved by the forward algorithm which is fully presented in [19]. As we use the health data to train the model, the likelihood probability can express how much the given sequence deviated from the health condition. To that end, likelihood probability can be used for degradation indicator. Since the extracted data are continuous and multidimensional, Gaussian mixture model of the continuous observation matrix in CHMM can be used to describe the distribution of these complicated data. Thus the CHMM-based indicator can not only capture the characteristics of the

multiple features but also decrease the stochastic influence in time domain. In general, log likelihood probability (LLP) values are smaller than zero. In order to improve its intelligibility, negative LLP (NLLP) is used as a health quantization indication in this study.

The Principle of the SVR-Based Prediction

Support vector machine (SVM) that is used for classification and regression analysis is a supervised learning model with associated learning algorithms. A version of SVM for regression is called support vector regression. The model produced by support vector classification (as described above) depends only on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin. Analogously, the model produced by SVR depends only on a subset of the training data, because the cost function for building the model ignores any training data which are close to the model prediction. Thus the basic function for SVR is given by

$$y = f(x) = \omega \phi(x) + b, \quad (\text{Eq 7})$$

where ω and b are the coefficients with target value y , and the $\phi(x)$ is a non-linear mapping function that can map the input vector x into high-dimension space for linear regression so that the non-linear SVR can be achieved. The estimation of these coefficients means solving

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2} \|\omega\|_2 \\ &\text{subject to} \quad \begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \varepsilon, \\ \langle \omega, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \end{aligned} \quad (\text{Eq 8})$$

where ε is a free parameter that serves as a threshold; all predictions have to be within an ε range of the true predictions. Slack variables are usually added into the above model to allow for errors and approximation in the case that the above problem is infeasible. More details about the estimation of the coefficients can be found in Ref. [20].

The main principle of SVR-based prediction is the iterated multi-step life prediction strategy. When n -point time series $X = \{x_1, x_2, \dots, x_n\}$ is given, we can obtain a training group expressed by $Y = \{y_1, y_2, \dots, y_m\}$, where $y_m = \{(x_m, x_{1+m}, \dots, x_{l+m}), x_{l+m+1}\}$, $(x_m, x_{1+m}, \dots, x_{l+m})$ is the input, and x_{l+m+1} is the output. When the SVR trained by the training group is ready, the prediction can be implemented using formula (7). The basic idea of iterated prediction is to use the last predicted result as the component of the input vector for the next prediction so that the future indicator can be captured. Then the remaining life can be computed according to the predicted steps and failure threshold.

The Principle of the Prediction Model

The historical data-based model can get the whole trend information of the full life cycle but less real-time performance. Even though the online data can obtain the trend information of real-time data, the long-term prediction accuracy is low. Therefore, a robust prediction model should contain the information of historical data and online data. For that purpose, two SVR prediction models based on online data and historical data, respectively, are used for residual life prediction, and two predicted results are obtained. Then the ultimate prediction result is captured by taking a weighted average of the two results. The formula of the residual life prediction is as follows:

$$RUL = \beta \cdot l_1 + (1 - \beta) \cdot l_2, \tag{Eq 9}$$

where β is the weight of the SVR model based on historical data with $\beta \in [0, 1]$, l_1 denotes the remaining life predicted by the historical data-based SVR model, and l_2 stands for the remaining life predicted by the online data-based SVR model.

If β is directly used for prediction, the real-time performance of prediction is very poor for the reason that β is constant. In addition, the accuracy of the two SVR prediction models is different in different degradation states. Then LMS is used for the adjustment of the weight β according to the predicted value and the real value, so that the prediction accuracy can be improved by the dynamic weights. LMS is a self-adaptive filtering algorithm. The formula of weight adjustment is given by

$$\beta_t = \beta_{t-1} + \mu_t \cdot e_t \cdot x_t, \tag{Eq 10}$$

where β_t is the weight at moment t , x_t is the indicator value at moment t , μ_t is the step length with $0 < \mu_t < 1/x_t$, and e_t is the relative prediction error of the two models. The estimation of e_t is achieved by the predicted indicator value and the actual value. Assume that λ_t^1 and λ_t^2 are, respectively, the predicted indicator value of the two SVR models at moment t , and λ_t is the actual value at moment t . Thus the computation of e_t is given by

$$e_t = 0.1 \cdot \left(\frac{|\lambda_t^1 - \lambda_t|}{|\lambda_t^2 - \lambda_t|} - 1 \right). \tag{Eq 11}$$

Here, $|\cdot|$ denotes the Euclid distance calculation.

The principle of the adaptive prediction is first to extract the general features from the vibration signal and construct the CHMM-based indicator with health data of these features. Then the SVR models based on online data and historical data can be built. When the incipient fault occurred, the residual life prediction is implemented. Thus the ultimate prediction result is computed according to formula (9). The prediction flow chart is shown in Fig. 1.

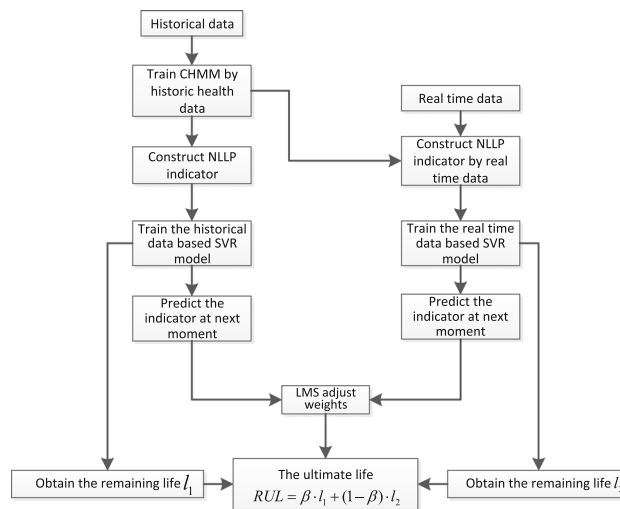


Fig. 1 The flow chart of the adaptive prediction method

Experimental Verification

The whole life cycle data of rolling element bearings are from NASA website. The experimental bearing type is Rexnord ZA-2115. The experimental rotating speed is 2000 rpm, and the sampling frequency is 20 kHz. Every sampling length of the data segment is 1 s, and the interval of the data segment is 10 min. Five groups of data are used in this paper: four of them treated as historical data are used to train the SVR model, while the last group is used for residual life prediction. The formula of the predicted error is given by

$$e = \frac{1}{n} \sum_{r=1}^n \frac{|x_r - d_r|}{d_r} \times 100\%, \tag{Eq 12}$$

where x_r stands for the predicted value at moment r and d_r stands for the actual value at moment r . Four kinds of general fault features are used for condition monitoring, namely kurtosis, root mean square value, peak-peak value, and peak value. The whole life cycle maps of the five groups of data are shown in Figs. 2, 3, 4, and 5.

The Construction of CHMM-Based NLLP Indicator

We firstly use the health data to train CHMM. As shown from Figs. 2, 3, 4, and 5, the health data can be determined from first data point to the 1000th data point. Then the CHMM-based NLLP indicator which is shown in Figs. 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15 can be obtained. As can be seen from Figs. 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15, the trend of the first 1000th data point is not easy to judge. However, the trend of CHMM-based NLLP indicator from 1500th data point to the end is obvious. Moreover, this significant performance can also be seen in Figs. 7, 9, 11,

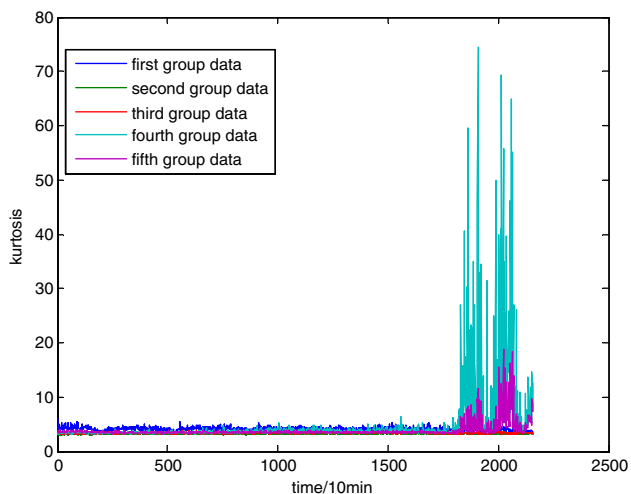


Fig. 2 The kurtosis map of the full life

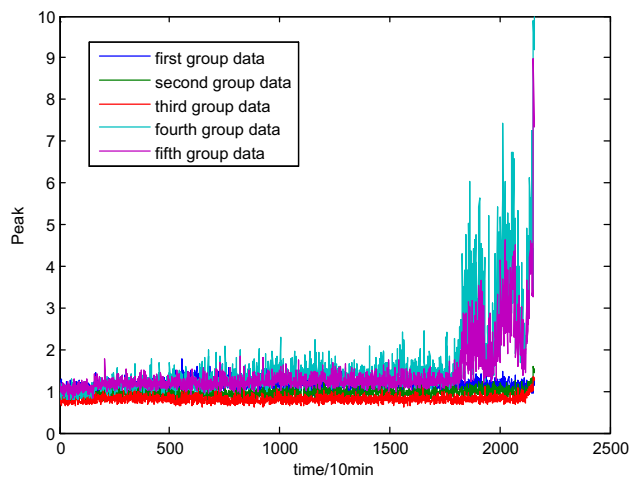


Fig. 5 The peak map of the full life

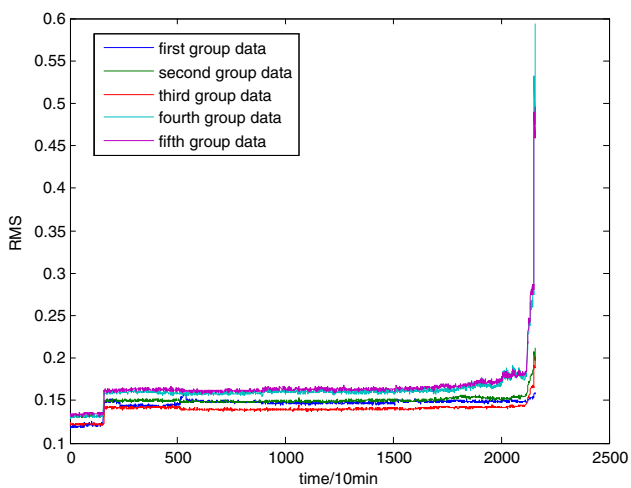


Fig. 3 The RMS map of the full life

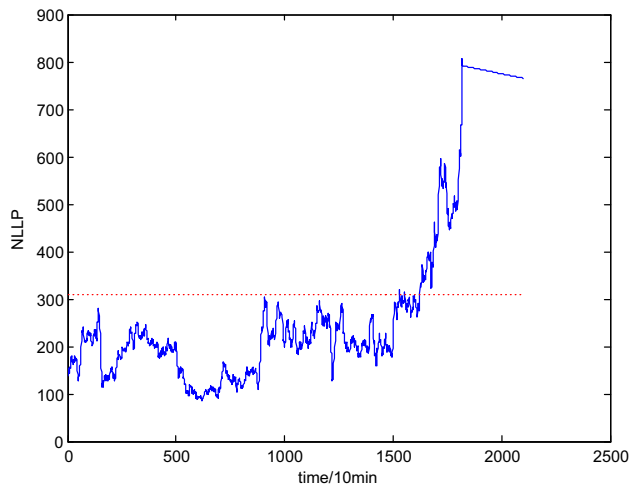


Fig. 6 The NLLP indicator of the first group of data

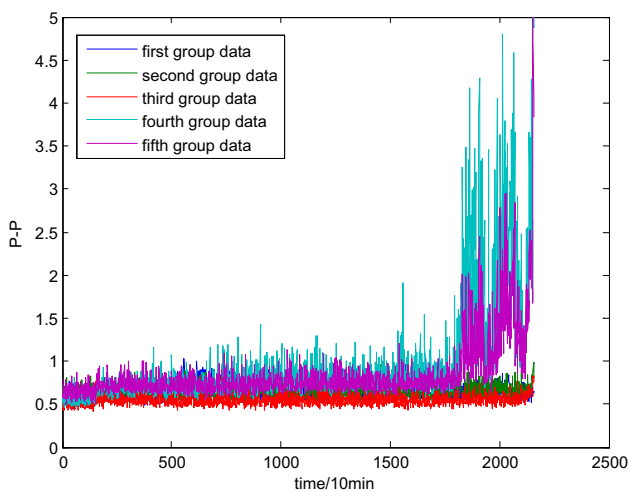


Fig. 4 The peak-to-peak map of the full life

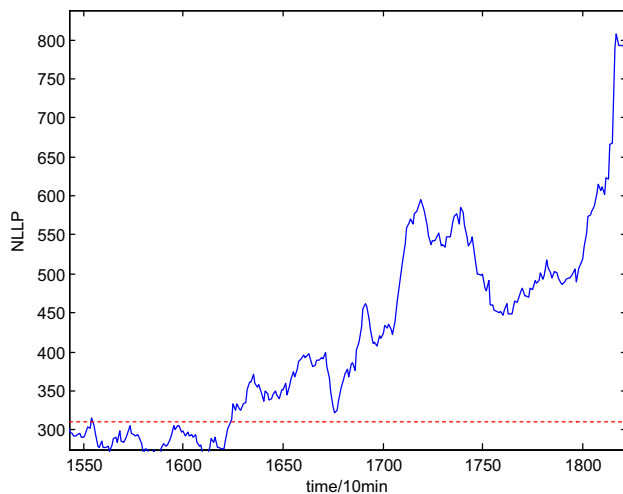


Fig. 7 Partial enlarged detail map of the NLLP indicator using the first group of data

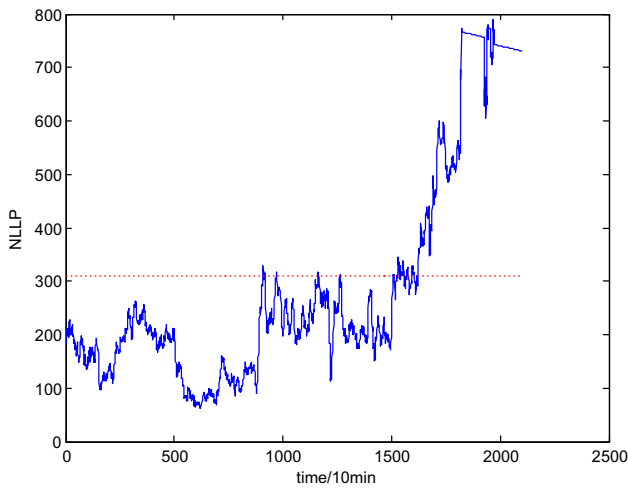


Fig. 8 The NLLP indicator of the second group of data

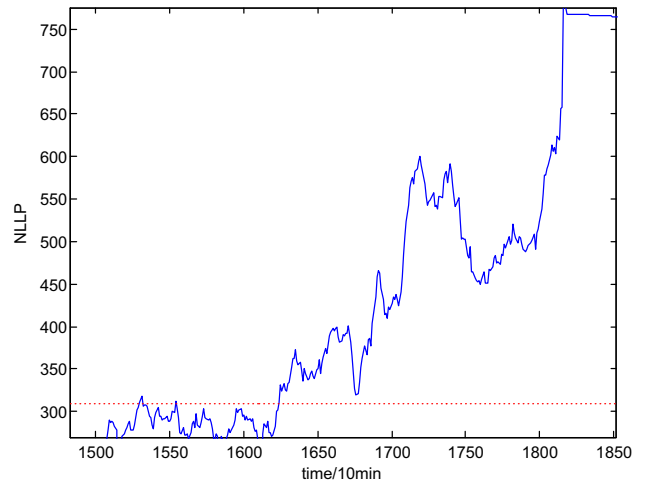


Fig. 11 Partial enlarged detail map of the NLLP indicator using the third group of data

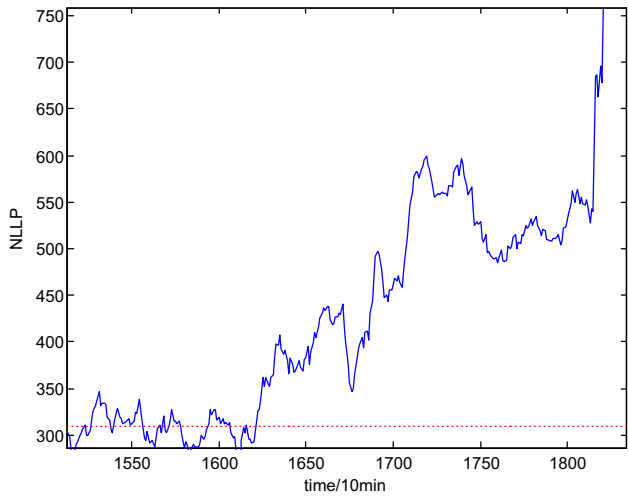


Fig. 9 Partial enlarged detail map of the NLLP indicator using the second group of data

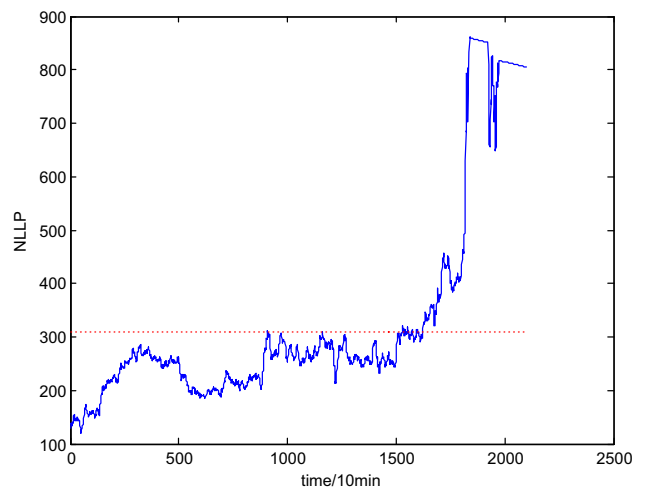


Fig. 12 The NLLP indicator of the fourth group of data

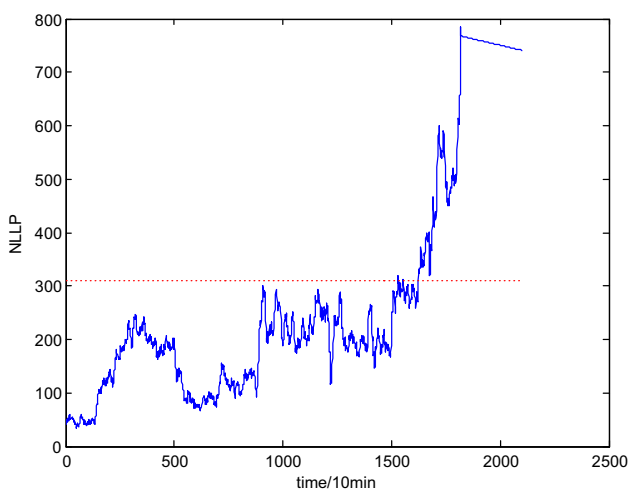


Fig. 10 The NLLP indicator of the third group of data

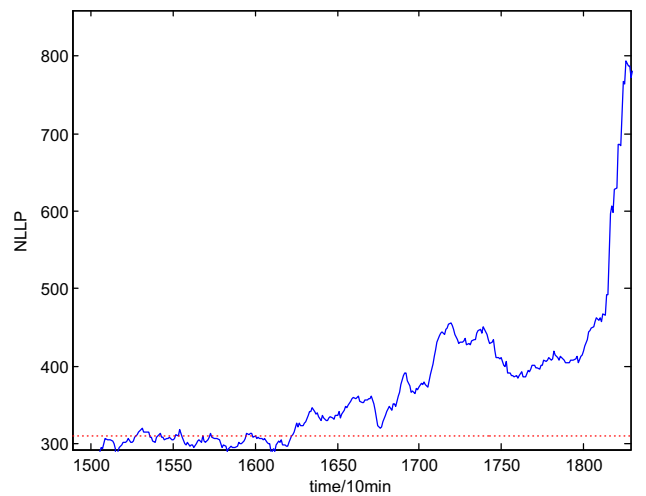


Fig. 13 Partial enlarged detail map of the NLLP indicator using the fourth group of data

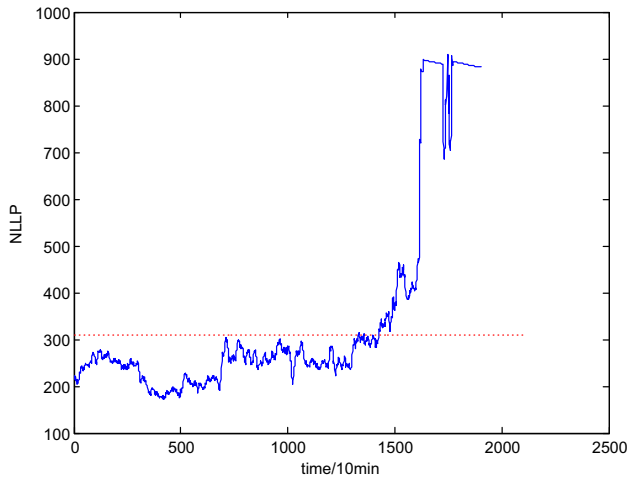


Fig. 14 The NLLP indicator of the fifth group of data

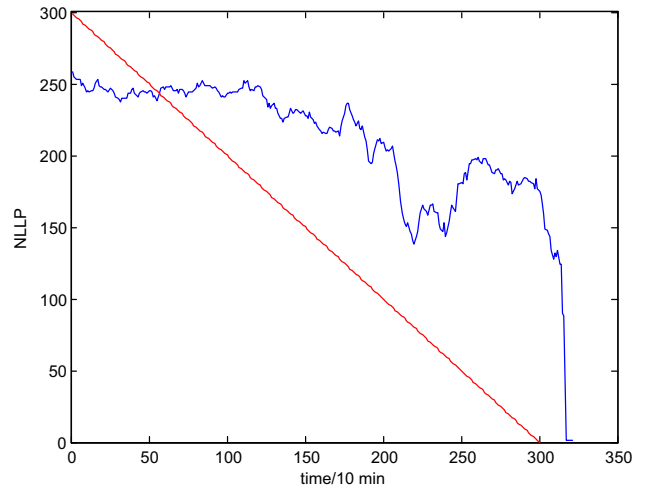


Fig. 17 The prediction map of the historical data-based SVR prediction method

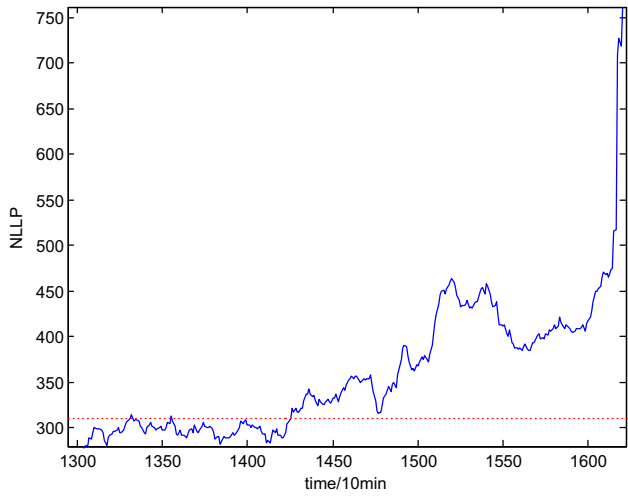


Fig. 15 Partial enlarged detail map of the NLLP indicator using the fifth group of data

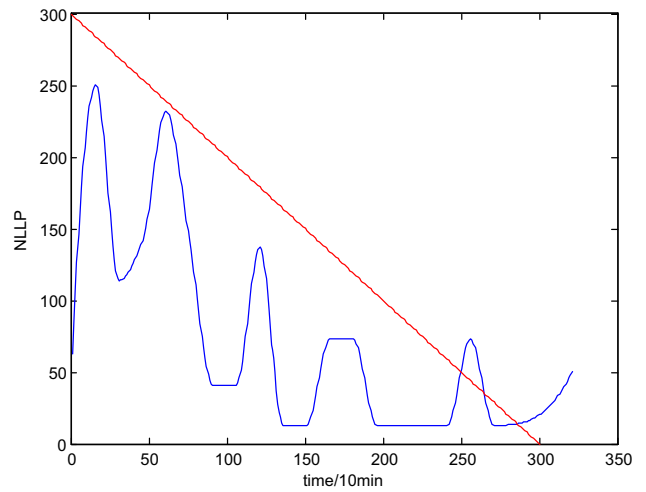


Fig. 18 The prediction map of the online data-based SVR prediction method

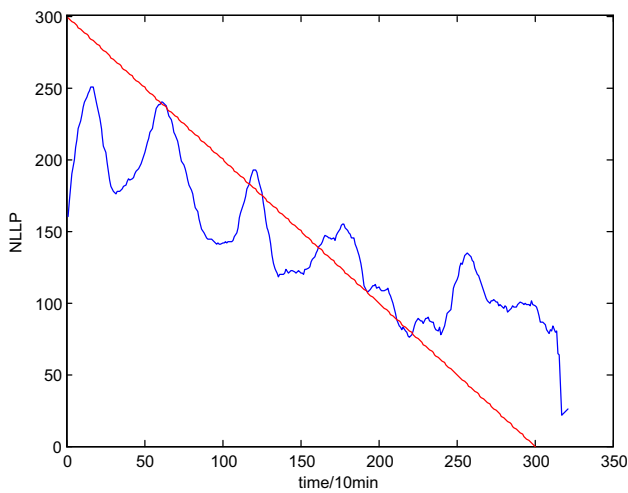


Fig. 16 The prediction map of the proposed prediction method

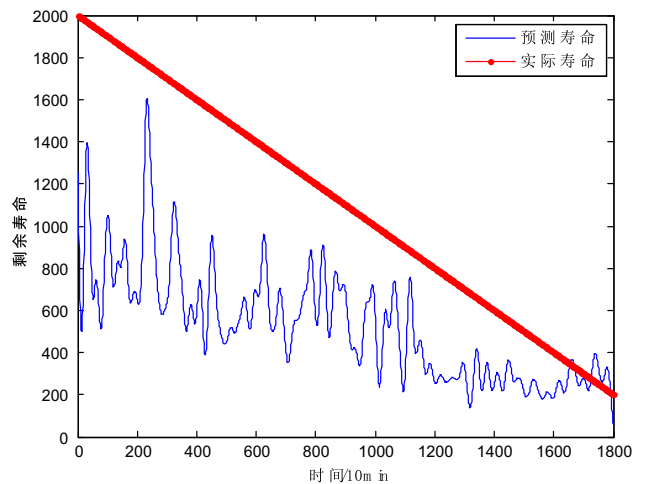


Fig. 19 The prediction map of the method in [2]

13, and 15. But in Figs. 3, 4, 5, and 6, a sudden increase in trend shows up instead of significant degradation trend. Otherwise, the trend of the proposed indicator is smoother and shows less outlier. In these figures of CHMM-based NLLP indicator, we can also choose the incipient fault threshold according to the red line. Therefore, the prediction can be implemented when NLLP indicator reaches the incipient fault threshold of about 300.

The Adaptive Prediction

In order to verify the validation of the proposed method, three other methods that are historical data-based SVR prediction method, online data-based SVR prediction method, and prediction method of [2] are used for comparison. Four groups of data treated as the historical data are used for the training of the historical data-based SVR model, while the last group of data treated as the online data is used for the training of the online data-based SVR model and monitoring. Figures 16, 17, 18, and 19, respectively, show the adaptive prediction map, the prediction map of the prediction model based on historical data, the prediction map of the prediction model based on real-time data, and the prediction map of the algorithm of Ref. [2], and the predicted errors are 0.251, 0.371, 0.382, and 0.384. In these figures, we can see that the predicted map of the proposed method is nearest to the real-life curve in comparison to other curves. Therefore, the experimental results show that the proposed method is effective.

Conclusion

In order to improve the prediction accuracy, a new method is proposed. Firstly, the CHMM is used for the construction of the degradation indicator that can show the significant incipient fault and the trend of the degradation. Then SVR and LMS are used for the construction of the adaptive prediction model that is composed of the historical data-based SVR model and the online data-based SVR model. The experimental results show that the historical data-based model or the online data-based model cannot provide high accuracy, while the proposed model based on the historical and online data can show great precision in comparison to other methods.

Acknowledgments The work described in this paper was supported by a grant from the National Defence Researching Fund (No. 9140A27020413JB11076).

References

1. H. Qiu, J. Lee, J. Lin et al., Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Adv. Eng. Inform.* **17**(3), 127–140 (2003)
2. J. Yan, C. Guo, X. Wang, A dynamic multi-scale Markov model based methodology for remaining life prediction. *Mech. Syst. Signal Process.* **25**(4), 1364–1376 (2011)
3. J. Yu, Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models. *Mech. Syst. Signal Process.* **25**(7), 2573–2588 (2011)
4. J.B. Yu, Bearing performance degradation assessment using locality preserving projections. *Expert Syst. Appl.* **38**(6), 7440–7450 (2011)
5. L. Liao, Discovering prognostic features using genetic programming in remaining useful life prediction. *IEEE Trans. Ind. Electron.* **61**(5), 2464–2472 (2014)
6. C. Chen, G. Vachtsevanos, M.E. Orchard, Machine residual useful life prediction: An integrated adaptive neuro-fuzzy and high-order particle filtering approach. *Mech. Syst. Signal Process.* **28**, 597–607 (2012)
7. Z. Tian, An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring. *J. Intell. Manuf.* **23**(2), 227–237 (2012)
8. P.Y. Chao, Y.D. Hwang, An improved neural network model for the prediction of cutting tool life. *J. Intell. Manuf.* **8**(2), 107–115 (1997)
9. N.Z. Gebraeel, M.A. Lawley, R. Li et al., Residual-life distributions from component degradation signals: a Bayesian approach. *IIE Trans.* **37**(6), 543–557 (2005)
10. J.P. Kharoufeh, S.M. Cox, Stochastic models for degradation-based reliability. *IIE Trans.* **37**(6), 533–542 (2005)
11. J.P. Kharoufeh, D.G. Mixon, On a Markov-modulated shock and wear process. *Naval Res. Logist. (NRL)* **56**(6), 563–576 (2009)
12. Orchard M E. AParticle filtering-based framework for on-line fault diagnosis and failure prognosis. Georgia Institute of Technology, 2007
13. P. Baraldi, M. Compare, S. Saucio et al., Ensemble neural network-based particle filtering for prognostics. *Mech. Syst. Signal Process.* **41**(1), 288–300 (2013)
14. M.S. Arulampalam, S. Maskell, N. Gordon et al., A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Sig. Process.* **50**(2), 174–188 (2002)
15. J.M. Lee, S.J. Kim, Y. Hwang et al., Diagnosis of mechanical fault signals using continuous hidden Markov model. *J. Sound Vib.* **276**(3), 1065–1080 (2004)
16. S. Zhou, J. Zhang, S. Wang, Fault diagnosis in industrial processes using principal component analysis and hidden Markov model[C]//American Control Conference. Proc. 2004 IEEE **2004**(6), 5680–5685 (2004)
17. B.S. Yang, A. Widodo, Support vector machine for machine fault diagnosis and prognosis. *J. Syst. Des. Dyn.* **2**, 12–23 (2008)
18. T. Benkedjough, K. Medjaher, N. Zerhouni et al., Remaining useful life estimation based on nonlinear feature reduction and support vector regression. *Eng. Appl. Artif. Intell.* **26**(7), 1751–1760 (2013)
19. R.I.A. Davis, B.C. Lovell, Comparing and evaluating HMM ensemble training algorithms using train and test and condition number criteria. *Formal Pattern Anal. Appl.* **6**(4), 327–335 (2004)
20. D. Basak, S. Pal, D.C. Patranabis, Support vector regression. *Neural Inf. Process. Lett. Rev.* **11**(10), 203–224 (2007)