

Remaining Useful Life Prediction of Rolling Bearings Based on Recurrent Neural Network

Yimeng Zhai¹, Aidong Deng^{1,*}, Jing Li^{1,2}, Qiang Cheng¹ and Wei Ren³

Abstract: In order to acquire the degradation state of rolling bearings and achieve predictive maintenance, this paper proposed a novel Remaining Useful Life (RUL) prediction of rolling bearings based on Long Short Term Memory (LSTM) neural network. The method is divided into two parts: feature extraction and RUL prediction. Firstly, a large number of features are extracted from the original vibration signal. After correlation analysis, the features that can better reflect the degradation trend of rolling bearings are selected as input of prediction model. In the part of RUL prediction, LSTM that making full use of the network's memory in time is used to improve the accuracy of RUL prediction. The proposed method is validated by life cycle experimental data of bearings, and the RUL prediction results of LSTM model are compared with Support Vector Regression (SVR) and Light Gradient Boosting Machine (LightGBM) models respectively. The results show that the proposed method is more suitable for RUL prediction of rolling bearings.

Keywords: Vibration signal, rolling bearing, RUL, LSTM neural network.

1 Introduction

Rolling bearings are the key parts to support and transfer torque, which have prompted its extensive use in rotating machinery, such as bearings in Wind Turbine Drive Train. According to statistics, 30% of the rotating machinery failures are caused by rolling bearings [Yu (2001)]. The reason is that the most of rolling bearings run in a harsh environment. Take rolling bearings in wind turbines for example, due to the intermittent and fluctuation of wind energy, the loads on the bearings have a strong time-varying and impact, resulting in high failure rate. Therefore, it is particularly important to predict the RUL of rolling bearings. Degradation state assessment and RUL prediction can not only effectively prevent sudden failure of mechanical equipment, but also maximize the use of the working capacity of key components, reduce maintenance costs and reduce unnecessary waste of resources. In recent years, it has become a research hotspot.

¹ National Engineering Research Center of Turbo-generator Vibration, Southeast University, Nanjing, 210096, China.

² School of Information Engineering, Nanjing Audit University, Nanjing, 211815, China.

³ Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong, 999077, China

*Corresponding Author: Aidong Deng. Email: dnh@seu.edu.cn.

RUL prediction is an important part of degradation analysis and failure prediction. The key issues are health feature extraction, degradation analysis and RUL prediction. Limited by the difficulty of dynamic model modeling [Wang (2018)], RUL prediction based on dynamic model has been unable to make a breakthrough. Data-driven method, which relies on the operation data collected by sensors, has a wide range of applications. Because there is no essential to further study the failure mechanism and accumulate a large number of expert experience. NASA proposed two strategies for RUL prediction based data-driven method: one is mapping n-dimensional features to 1-dimensional health indicator, and then using curve fitting and extrapolation to predict; the other is constructing degenerate feature vectors containing n-dimensional data, giving the residual life of the target, which can help achieving pattern matching [Liu and Li (2017)]. Considering that the health indicators obtained by feature fusion have no clear physical meaning, the second strategy is adopted in this paper. The data-driven RUL prediction methods that are commonly used include regression model, Bayesian reasoning [Zhao, Jiang and Long (2018)], Gauss mixture model [Zhang, Kang and Zhao (2014)] and other statistical analysis methods, as well as artificial intelligence methods such as Fuzzy Decision Tree, Artificial Neural Network (ANN) [Gebraeel, Lawley, Liu et al. (2004); Huang, Xi, Li et al. (2007)], SVR [Lei, Chen, Li et al. (2016); Loutas, Roulias and Georgoulas. (2013); Tse and Shen (2015)] and Hidden Markov Model (HMM) [He and Wang (2014)].

In recent years, with the continuous development of deep learning technology, some deep learning models have been gradually applied [Zhao, Wu, Zhang et al. (2018)]. Deep learning model is a kind of deep neural network model with multiple levels of non-linear mapping, which can abstract input signal layer by layer and extract features, and excavate deeper potential laws. In many deep learning models, Recurrent Neural Network (RNN) introduces the concept of time series into network structure design, which makes it more adaptable in time series data analysis. In order to solve the problems of gradient disappearance and gradient explosion in RNN, LSTM, Gated Recurrent Unit (GRU) and Bi-directional Recurrent Neural Network (BRNN) which are all variants of RNN are widely used.

The aim of RUL prediction is to obtain the deterioration trend of bearings at the current moment. As a result, it is necessary to have a certain memory of features of the historical moment. Therefore, this paper presents a method of rolling bearing RUL prediction based on LSTM, which mainly includes feature extraction and RUL prediction. Firstly, the feature vectors of the original vibration signal are extracted, including variance, root mean square value, kurtosis, skewness, peak value, entropy and wavelet coefficients. After correlation analysis, the feature vectors which can better reflect the trend of bearing degradation are selected to be input of the prediction model. In the part of RUL prediction, some samples are input into LSTM network in batches as training sets to complete model training and network parameter adjustment. After the model is constructed, the model is tested with test sets, and the RUL prediction values of test sets are obtained. In addition to LSTM network model, SVR and LightGBM methods are also used to compare and prove the effectiveness of the proposed method.

2 Long short term memory

Different from the ANN, there are some ring structures existing in RNN. As shown in the Fig. 1, the output state \hat{y}_t of the network is not only related to the input x_t , but also to the network state h_{t-1} . The mathematical expression is as follows (1) (2). RNN extends the time dimension on the basis of space, which can make correct prediction by using the correlation information of sequence data. With the increase of time interval, the gradient norm of back propagation parameters decreases exponentially, which easily leads to the disappearance of gradient.

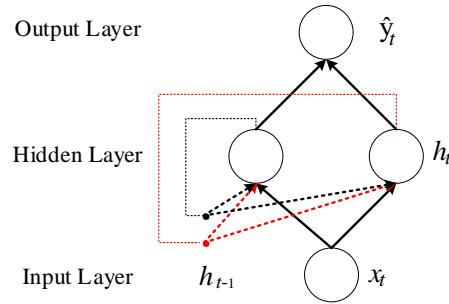


Figure 1: The network structure of RNN

$$h_t = g(w_{xh}x_t + w_{hh}h_{t-1} + b_h) \tag{1}$$

$$\hat{y}_t = g(w_{hy}h_t + b_y) \tag{2}$$

Where x_t is the input of the input layer at time t ; \hat{y}_t is the output of the output layer at time t . h_{t-1} is the output of the hidden layer at time $t-1$. h_t is the output of the hidden layer at time t . w_{xh} is weight values between an input layer x_t and a hidden layer h_t at time t . w_{hh} is weight values between an hidden layer h_{t-1} at time $t-1$ and a hidden layer h_t at time t . w_{hy} is weight values between an hidden layer h_t at time t and a output layer \hat{y}_t at time t . b_h and b_y are bias of hidden layer and output layer, respectively.

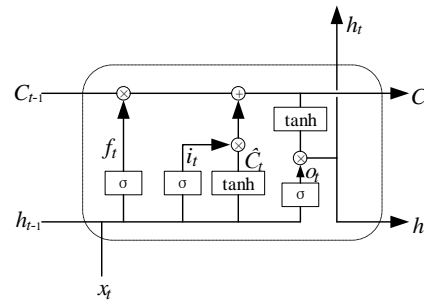


Figure 2: The structure of LSTM memory unit

LSTM is a variant of RNN. RNN can only have short-term memory due to the disappearance of gradient. However, LSTM network combines short-term memory with long-term memory by introducing “gate” structure, which can increase or forget information to the cell state. As a result, information can selectively pass through, and

solve the problem of gradient disappearance in a manner.

As shown in Fig. 2, each LSTM memory unit contains input gate, forget gate and output gate. First, the forget gate outputs the number between 0 and 1 through sigmoid layer (forgetting threshold). The output multiplies the cell state of the previous moment to control information forgetting. Then, the pointwise multiplication of the input threshold layer (sigmoid layer) and the temporary cell state is output through input gate to control information input, and complete the update of cell state. Last, the cell state was processed by tanh layer, and the output of cell state was controlled by sigmoid layer. The output equations of each gate are as follows:

$$\begin{cases} f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + b_f) \\ i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + b_i) \\ \hat{C}_t = \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c) \\ C_t = f_t * C_{t-1} + i_t * \hat{C}_t \\ o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + b_o) \\ h_t = o_t * \tanh(C_t) \end{cases} \quad (3)$$

where x_t is the input of LSTM cell unit at time t . h_{t-1} is the output of LSTM cell unit at time $t-1$. h_t is the output of LSTM cell unit at time t . C_{t-1} is the cell state at time $t-1$. C_t is the cell state at time t . \hat{C}_t is the temporary cell state at time t . i_t, f_t and o_t are output values of the input node, the input gate, the forget gate and the output gate, respectively. b_i, b_f, b_o and b_c are bias of the input node, the input gate, the forget gate and the output gate, respectively. w_{xi}, w_{xf}, w_{xo} and w_{xc} are weight values between an input layer x_t and a hidden layer h_t at time t , respectively. w_{hi}, w_{hf}, w_{ho} and w_{hc} are hidden layer weight values between time t and $t-1$, respectively.

3 Experiments and result analysis

The life cycle data of rolling bearings used in this paper are derived from IEEE PHM 2012 Data Challenge [Nectoux, Gouriveau, Medjaher et al. (2012)]. The experimental data are from the PRONOSTIA test rig, and the structure is shown in Fig. 3. By adding additional load or increasing speed to bearings, the purpose of accelerated failure can be achieved. The experimental bearings are all running at 1800 r/min speed and 4000N load. Acceleration sensors collected data every 10 seconds. The time length of each data acquisition is 0.1 seconds, that is to say, 2560 data points are collected each time. When the acceleration amplitude exceeds 20g, it is considered invalid, and the experiment is finished. Seven bearings under this working condition are running from normal state to failure. The number of bearing samples are shown in Tab. 1.

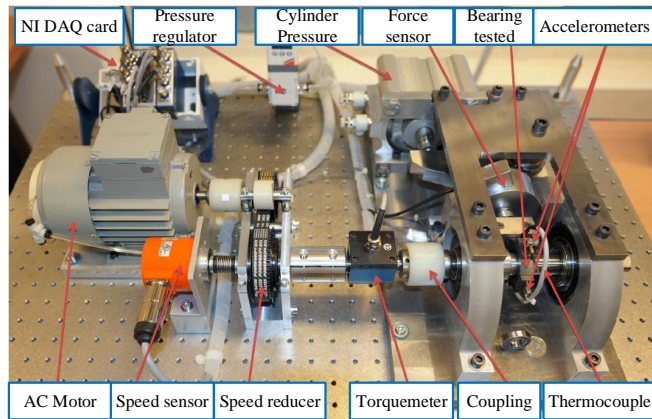


Figure 3: The structure of PRONOSTIA test rig

Table 1: Number of bearing samples

	No.1	No.2	No.3	No.4	No.5	No.6	No.7
Number of samples	2801	871	2375	1428	2463	2448	2259

As shown in Fig. 4, a full life cycle waveform of No. 1 bearing is presented. It can be seen that in the early stage of bearing running, the vibration amplitude fluctuates smoothly, then the fluctuation buoyancy increases. At last, the amplitude increases sharply in the final stage. The whole degradation trend accords with the degradation process of equipment operation.

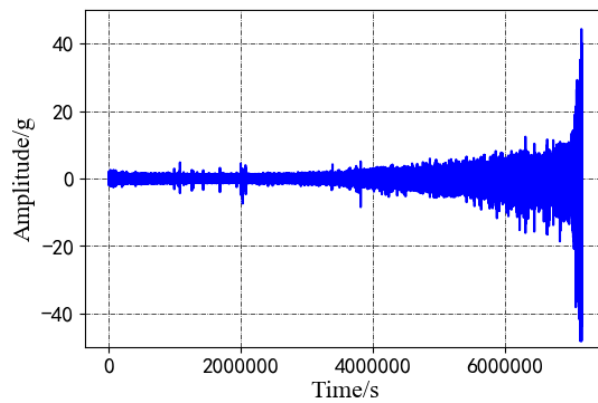


Figure 4: The full life cycle waveform of No. 1 bearing

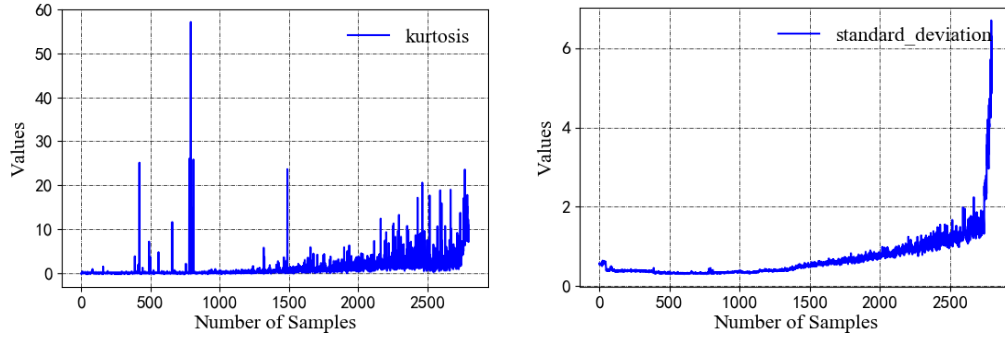


Figure 5: Feature sequence diagram of standard deviation and kurtosis for No. 1 bearing

3.1 Feature extraction

Taking samples as units, feature extraction is carried out one by one, i.e., extracting corresponding features from each sample, and then combining the features from each sample into a new feature sequence. The extracted features include variance, root mean square value, kurtosis, skewness, peak value, entropy, wavelet coefficients, FFT coefficients and so on. As shown in Fig. 5, there are feature sequence diagrams of standard deviation and kurtosis for No.1 bearing. Compared with the waveform of bearing life cycle, it is found that the standard deviation is more consistent with the degradation trend of bearing. Because the kurtosis characteristic whose feature values at early stage even exceed the values in later stage is more sensitive to impact, burrs can be observed in the initial stage of bearing operation, which will affect the RUL prediction accuracy.

Therefore, based on the extracted features, feature selection is carried out, and correlation analysis is used to calculate the correlation between feature series and time series, then the trend indicator of feature is obtained. For a feature sequence $F=[g(t_1), g(t_2), \dots, g(t_K)]$ and time sequence $T=[t_1, t_2, \dots, t_K]$, the trend indicator $Corr(F, T)$ can be get through formula (4). Where $g(t_k)$ is feature value at time t_k , K is the length of sample time.

$$Corr(F, T) = \frac{\left| K \sum_k g(t_k) t_k - \sum_k g(t_k) \sum_k t_k \right|}{\sqrt{\left[K \sum_k g(t_k)^2 - \left(\sum_k g(t_k) \right)^2 \right] \left[K \sum_k t_k^2 - \left(\sum_k t_k \right)^2 \right]}} \quad (4)$$

3.2 RUL prediction

In addition to No. 3 bearing, the full life cycle samples of remaining bearings are taken as training sets to construct the RUL prediction model. The samples of No. 3 bearing are taken as test sets, and the accuracy of prediction results is characterized by Root Mean Square Error (RMSE). The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2} \quad (5)$$

Where y_n is the real RUL of n th samples of bearing; \hat{y}_n is the predicted RUL of n th samples of bearing; N is the number of samples.

For training sets $\{x_n, y_n\}_{n=1}^N$, where $x_n \in R^{m \times 1}$ is selected m features of n th, $y_n \in [0, 1]$ is its associated label which indicates normalized RUL of n th. Except for building LSTM network prediction model, SVR and LightGBM models are used to compare. There is the comparison of the actual and predicted values of these three methods in Fig. 6. LSTM network has 3 layers, 30 hidden layer units, *learning rate* is 0.1, *time step* is 30, *batch_size* is 100, and iteration step is 20,000. In SVR model, penalty parameter $C=32$, $g=1$. The parameters settings of LightGBM are: the *learning rate* $\eta=0.1$, the *num_leaves* $\alpha=100$, the *max_depth* $h=12$ and the number of trees is $n_estimators=3000$.

As can be seen from the Fig. 6: (1) The results predicted by the three methods are all the same as the bearing degradation trend. (2) In general, the prediction results in the mid-term operation of bearings are better than those in the early and later stages. As we know from Fig. 5, most of the features have the characteristics of basically unchanged in the early stage, slow change in the middle stage and drastic change in the later stage. However, the RUL is the time interval between sample sampling time and the end of the experiment, i.e., it is a straight line with slope of -1. The change of vibration amplitude, which basically coincides with the trend of residual life change, is slow in the mid-operation period. Therefore, the three methods all get better prediction results in the intermediate stage. (3) The comparison of the three methods shows that the RUL prediction results based on LSTM network are the best, not only in the intermediate stage, but also in the early and later stages of bearing operation. The RMSE of LSTM network is 0.050, which is half of LightGBM, and one-third of SVR. Therefore, the method proposed in this paper can accurately predict the RUL of rolling bearings and provide a basis for predictive maintenance.

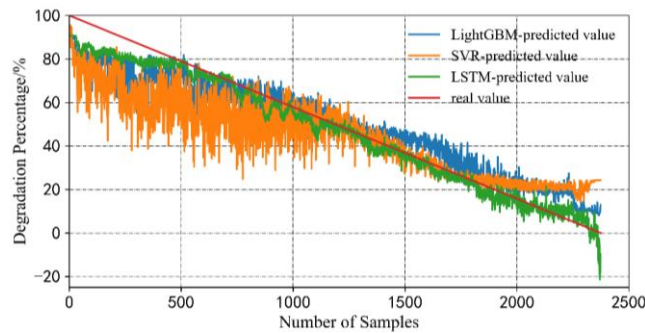


Figure 6: The comparison of the actual and predicted values of these three methods

4 Conclusion

In this paper, a novel method for predicting the RUL of rolling bearings based on LSTM is proposed. Firstly, feature selection is conducted through trend indicator, and then a feature vector reflecting the degradation trend of rolling bearings is constructed, which can be as input of LSTM prediction model. Making full use of the memory of LSTM

network for historical moment features can improve the accuracy of prediction. Compared with SVR and LightGBM, it is proved that LSTM network model can accurately predict the RUL of bearings in the whole life stage. Monitoring the degradation trend of rolling bearings and the stability of rolling bearings in the whole life cycle can not only ensure the ideal utilization rate, but also avoid major accidents, which has significant application value.

Acknowledgement: This work is supported by the National Nature Science Foundation of China (No. 51875100). The authors would like to thank anonymous reviewers and the associate editor, whose constructive comments help improve the presentation of this work.

References

- Chen, R.; Huang, X.; Yang, L.; Tang, B.; Yu, T. et al.** (2018): Rolling bearing fault identification based on convolution neural network and discrete wavelet transform. *Journal of Vibration Engineering*, vol. 31, no. 5, pp. 883-891.
- Chen, Z.; Li, C.; Sanchez, R. V.** (2015): Gearbox fault identification and classification with convolutional neural networks. *Shock and Vibration*, vol. 15, no. 8, pp. 18-28.
- Gebraeel, N.; Lawley, M.; Liu, R.; Parmeshwaran, V.** (2004): Residual life predictions from vibration-based degradation signals: a neural network approach. *IEEE Transactions on Industrial Electronics*, vol. 51, no. 3, pp. 694-700.
- He, Z.; Wang, S.** (2014): Remaining lifetime prediction based on time-varying state transition probabilities of Hidden Semi-Markov Model. *Journal of Hunan University (Natural Sciences)*, vol. 41, no. 8, pp. 47-53.
- Huang, R.; Xi, L.; Li, X.; Liu, C. R.** (2007): Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems & Signal Processing*, vol. 21, no. 1, pp. 193-207.
- Lei, Y.; Chen, W.; Li, N.; Lin, J.** (2016): A relevance vector machine prediction method based on adaptive multi-kernel combination and its application to remaining useful life prediction of machinery. *Journal of Mechanical Engineering*, vol. 52, no. 1, pp. 87-93.
- Liu, F. Li, Q.** (2017): Development and application of big data technology in aviation industry. *Telecommunication Engineering*, vol. 57, no. 7, pp. 849-854.
- Loutas, T. H.; Roulias, D.; Georgoulas, G.** (2013): Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression. *IEEE Transactions on Reliability*, vol. 62, no. 4, pp. 821-832.
- Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Chebelmorello, B. et al.** (2012): PRONOSTIA: an experimental platform for bearings accelerated degradation tests. *IEEE International Conference on Prognostics & Health Management*, pp. 1-8.
- Tse, P. W.; Shen, C. Q.** (2015): Remaining useful life estimation of slurry pumps using the health status probability estimation provided by support vector machine. *Engineering Asset Management-Systems, Professional Practices and Certification*, pp. 87-98.
- Wang, W.; Guo, H.; Fan, Q.** (2018): Predict remaining useful life of aerospace engine based on CNN and GRU. *Proceedings of the 37th Chinese Control Conference*.

Yang, Z.; Yao, Z.; Wang, R. (2017): Study on fault diagnosis of gearbox based on hmm. *Acta Energiae Solaris Sinica*, vol. 38, no. 9, pp. 2574-2581.

Yu, H. (2001): *Equipment Fault Diagnosis Engineering*. Metallurgical Industry Press.

Zhang, X.; Kang, J.; Zhao, J.; Xiao, L.; Cao, D. et al. (2014): Equipment degradation state identification and residual life prediction based on MoG-BBN. *Journal of Vibration and Shock*, vol. 33, no. 8, pp. 171-179.

Zhao, S.; Jiang, C.; Long, X. (2018): Remaining useful life estimation of mechanical systems based on the data-driven method and Bayesian Theory. *Journal of Mechanical Engineering*, vol. 54, no. 12, pp. 115-124.

Zhao, X.; Wu, J.; Zhang, Y.; Shi, Y.; Wang, L. (2018): Fault diagnosis of motor in frequency domain signal by stacked de-noising auto-encoder. *Computers, Materials & Continua*, vol. 57, no. 2, pp. 223-242.