

Remaining Useful Life Prediction of Machinery based on K-S Distance and LSTM Neural Network

Yang Ge^{a,b,*}, Lanzhong Guo^{a,b}, and Yan Dou^{a,b}

^aKey Construction Laboratory on Elevator Intelligent Safety of Jiangsu Province, Changshu, 215500, China

^bSchool of Mechanical Engineering, Changshu Institute of Technology, Changshu, 215500, China

Abstract

The remaining useful life is key to the decision-making of machinery maintenance. The online prediction of remaining useful life has become a very urgent need for mechanical equipment with high reliability requirements. The aim of this paper is to provide a simple and effective method for predicting the remaining life of the machine under the condition of small sample. The Kolmogorov-Smirnov test theory is used to extract the health state feature of the machine. Based on the Long and Short Term Memory (LSTM) theory, an online method of remaining useful life prediction is proposed. The bearing life vibration data verification shows that the Kolmogorov-Smirnov distance is sensitive to the development and expansion of the defects. Furthermore, the proposed method of remaining useful life prediction based on LSTM theory has high prediction accuracy. The technician can then use this method to take appropriate maintenance operations.

Keywords: remaining useful life prediction; kolmogorov-smirnov test; long and short term memory; maintenance strategy; small sample

(Submitted on October 23, 2018; Revised on November 25, 2018; Accepted on December 26, 2018)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

More and more online monitoring technology is used to monitor machinery with high reliability demands [1]. Before mechanical failure, in order to maintain higher reliability and lower operating costs, the machine is maintained at the appropriate time by using actual situation information [2]. This maintenance strategy is called Condition-Based Maintenance (CBM) or Prognostic and Health Management (PHM). It is designed to conduct maintenance before machinery breakdown occurs based on historical trend data and real-time evaluation of machinery conditions, including the operating environment, risk estimation, and remaining life prediction [3].

In general, the methods of remaining useful life prediction are divided into three categories: failure model-based [4-5], probabilistic statistics-based [6-7], and data driven-based [8-9]. The prediction based on the failure model can penetrate the essence of the object and have high precision, but it is more accurate for the mathematical model of the object. For complex electromechanical devices, it is often difficult to establish precise and complete mathematical models, which leads to the limited application of the model for degradation evaluation and prediction. The method based on probability statistics mainly establishes the probability model of each parameter variation and fault loss model based on the historical data of the object system. It compares the current multi-parameter probability state space with the established probability model to predict the service life. In comparison with the above two methods, the data driven-based life prediction overcomes the limitation of requiring accurate mathematical models. With the development of machine learning, the data driven method has become mainstream in the field of remaining life prediction, and it relies only on past observation data and statistical models [10-11].

Normally, the first step of the data driven-based remaining life prediction process is to extract useful features from the original signals. The feature should reflect the health state of device. In practice, it is difficult to determine the best features, because many features are involved in certain stages but cannot keep the trend throughout the life cycle. For example, the root mean square (RMS) of the vibration signal is a feature commonly adopted for life prediction. According to literature

* Corresponding author.

E-mail address: geyang@cslg.edu.cn

[12], it shows an increasing trend at the beginning of fault growth but decreases later, and there is no obvious sign of failure. The time domain, frequency domain, and time–frequency domain features of the original signals are often used for life prediction [13–14], but it is difficult to determine which features are more sensitive to defect development and the propagation process. Regardless of feature extraction or selection, the calculation is normally very complex.

After appropriate features are selected, the next step of the data-driven based RUL prediction process is to establish a prediction model. More and more artificial intelligence algorithms have been applied to predict the RUL, such as support vector machine [15–17], extreme learning machine [18], joint approximate diagonalization of eigenmatrices (JADE) algorithm [14], particle filter [19], deep learning [20], and so on. These models can estimate health states and update model parameters according to real-time information. However, except for the prediction model, the prediction accuracy is closely related to the selected features and sample size. Therefore, how to establish a RUL prediction model under small sample conditions and improve the prediction accuracy is a practical significant problem that needs to be studied and solved.

The rest of the paper is organized as follows. Section 2 introduces the basic theory of the Kolmogorov-Smirnov test and Long and Short Term Memory. Section 3 shows the proposed prediction method. In Section 4, the method is demonstrated using data of accelerated degradation tests on rolling bearings. Conclusions are drawn in Section 5.

2. Basic Theory of K-S Test and LSTM

2.1. Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test is a non-parametric statistical method used to detect the similarity between two random distributions. It calculates the maximum vertical distance between the empirical distribution function of the measured signal and the reference signal. The maximum vertical distance can be used as a measure of the similarity between empirical distribution functions. According to the similarity between the measured signal and the reference signal, we can determine whether the state of machinery has been transferred.

Assume the sample data $X = \{x_1, x_2, \dots, x_n\}$, and the observed values x_1, x_2, \dots, x_n of the sample can be arranged as $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ in increasing order. The empirical distribution function (EDF) $F_X(z)$ of the sample is defined as Equation (1).

$$F_X(z) = \begin{cases} 0, & z < x_{(1)} \\ \frac{n}{N}, & x_{(n)} < z < x_{(n+1)}, \quad n = 1, 2, \dots, N-1 \\ 1, & z \geq x_{(N)} \end{cases} \quad (1)$$

Assume $R_X(z)$ is EDF of the reference random signal sample and $F_X(z)$ is EDF of the measured random signal sample. The Kolmogorov-Smirnov distance between $R_X(z)$ and $F_X(z)$ is defined as Equation (2).

$$D = \sup_{-\infty < z < \infty} |F_X(z) - R_X(z)| \quad (2)$$

2.2. Prediction Method based on LSTM

Long and Short Term Memory (LSTM) is a special Recurrent Neural Network (RNN). The classical RNN model is shown in Figure 1. For a given sequence $x = (x_1, x_2, \dots, x_n)$, using the RNN model, a prediction sequence $y = (y_1, y_2, \dots, y_n)$ can be obtained by iterative calculation of Equations (3) and (4).

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (3)$$

$$y_t = W_{hy}h_t + b_y \quad (4)$$

Where $h = (h_1, h_2, \dots, h_n)$ is the hidden sequence, W_{xh} is the weight coefficient matrix of input layer to hidden layer, b_h is the offset vector of the hidden layer, b_y is the offset vector of the output layer, and f denotes the activation function, which is generally non-linear, such as tanh or ReLU.

The classical RNN (simple-RNN) is equivalent to some multi-layer Deep Neural Networks (DNNs) spread on time series. This model is prone to gradient extinction or gradient explosion [21–22]. LSTM model can learn long-term

dependency information while avoiding the problem of gradient extinction. LSTM added a memory unit to record historical information in the neural node of the RNN hidden layer and added three kinds of gates (Input, Forget, Output) to control the use of historical information. The internal structure of neuron in LSTM is shown in Figure 2.

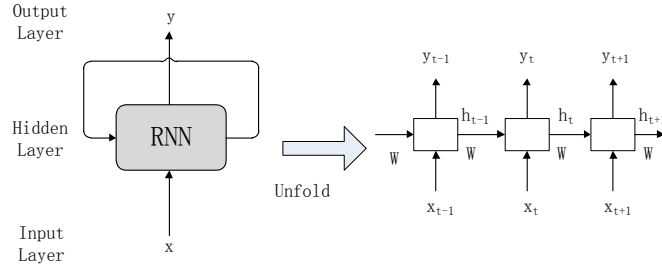


Figure 1. RNN model

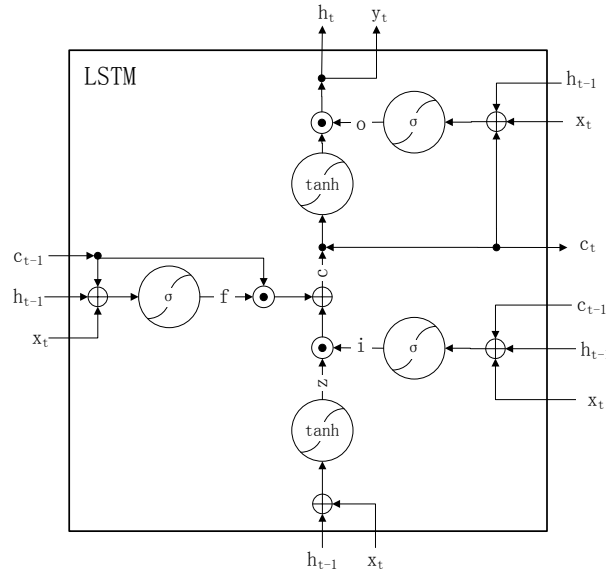


Figure 2. The internal structure of neuron in LSTM

Where i, f, c , and o respectively represent the input gate, the forgetting gate, the unit state, and the output gate; W and b are the corresponding weight coefficient matrix and the offset vector, respectively; and σ and \tanh respectively represent the sigmoid and hyperbolic tangent activation functions. The forward calculation formula is shown as Equations (5) to (9).

$$i_t = \text{sigmoid}(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (5)$$

$$f_t = \text{sigmoid}(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (6)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (7)$$

$$o_t = \text{sigmoid}(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (8)$$

$$h_t = o_t \tanh(c_t) \quad (9)$$

3. Remaining Useful Life Prediction Method of Machinery based on K-S Distance and LSTM

The general method to predict the remaining useful life of machinery is to first classify the operating status of the machinery, then construct the index accurately, describe the property degradation rule, determine the performance degradation start-stop time and critical failure time of the machinery, and finally establish a model to predict the remaining useful life based on this information. A remaining useful life prediction method of machinery based on K-S distance and LSTM presented in this paper is shown in Figure 3, and the main steps are as follows:

- (1) Collect the original signal from the measured machinery at equal time intervals, and the time interval is set as T .
- (2) Calculate the K-S distance of each collected sample as the evaluation index of the machinery performance degradation degree. If the calculation result is greater than the early degradation threshold, the machinery has entered a degraded state; if the calculation result is greater than the failure threshold, the machinery cannot continue to be used.
- (3) Continuously obtain N samples at the beginning of the early degradation point, calculate their K-S distance with the health state, and form the K-S distance sequence as $X = [KS^1, KS^2, \dots, KS^N]$.
- (4) Use the degradation index sequence $X = [KS^1, KS^2, \dots, KS^N]$, train the LSTM model according to Equations (5) to (9), and complete one step forward prediction based on this.
- (5) Determine whether the predicted K-S distance reaches the failure threshold. If not, add the predicted value to the degradation sequence, delete the first value of the original sequence, keep the sequence length n constant, update the degradation sequence dynamically, and repeat step (4) to complete P steps forward prediction.
- (6) If the predicted K-S distance reaches the failure threshold after P steps forward prediction, terminate the prediction process. At this point, the remaining useful life of machinery can be calculated according to the predicted steps P and the signal collection interval T by formula $RUL = P \times T$.

The LSTM model can be trained with a small number of degradation indexes. Using the trained LSTM model to predict trend changes of the degradation index, the time that the degradation index reaches the failure threshold can be estimated. In this way, the remaining useful life prediction of machinery under small sample conditions can be realized.

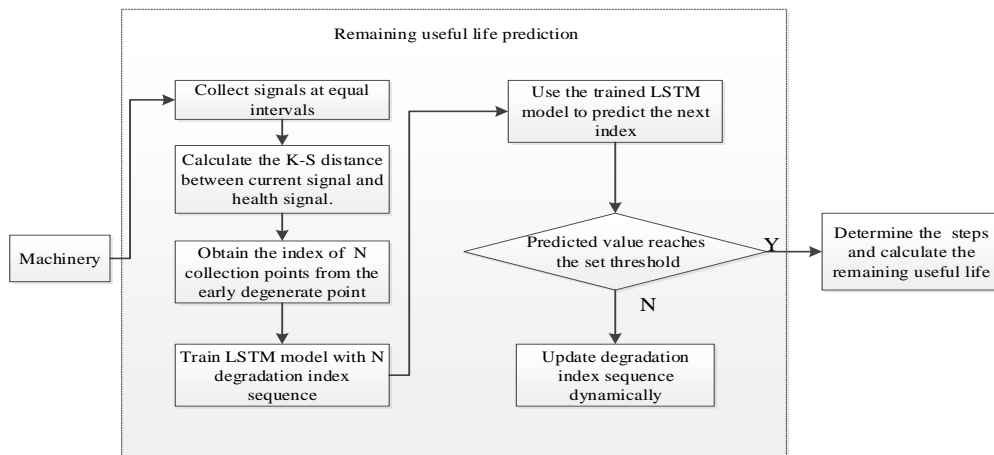


Figure 3. Remaining useful life prediction of machinery based on K-S distance and LSTM

4. Case Study

Data of bearing life vibration, from the Intelligent Maintenance System (IMS) of the University of Cincinnati, is used to verify the method proposed in this paper [23]. The whole life data set of bearing 1 is analyzed in this paper. The sampling frequency of bearing 1 is 20 kHz. Data was collected every 10 minutes, and 20480 data were collected every time. When bearing 1 continuously ran for 163.3 hours, the outer ring was damaged seriously, and 980 sets of data were collected.

4.1. Bearing Performance Degradation Evaluation

In engineering, the mean square root and kurtosis index of vibration data are generally used to monitor the operation status of machinery. The root mean square value and kurtosis of bearing 1 are shown in Figure 4. The initial state of the bearing can be regarded as a healthy state. Based on the K-S distance proposed in this paper, the bearing degradation evaluation index is shown in Figure 5. From Figures 4 and 5, the K-S distance index of the bearing's degenerative tendencies is approximately the same as the root mean square and kurtosis index. The bearing is in a healthy state at the beginning, and all three indexes are in a stable state. In the middle part, bearing performance begins to decline, and when bearing fails at the end, the changes of the three indexes are all drastic at the same collection point. However, compared with the mean square root and kurtosis of the bearing, the K-S distance is better than the mean square root and kurtosis in the description

of bearing degradation. From Figure 5, it can be seen clearly that the K-S distance index of the bearing increases continuously from the early degenerate point (the 533th collection point) to the failure point (the 962nd collection point). In contrast, the changing trend of root mean square and kurtosis is not obvious.

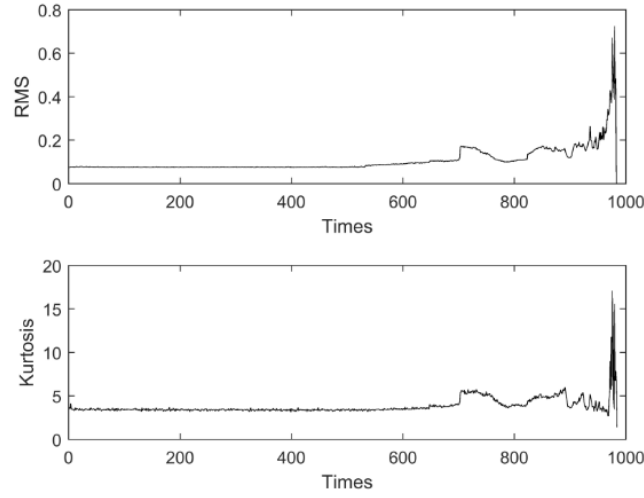


Figure 4. Mean square root and kurtosis of bearing vibration signal

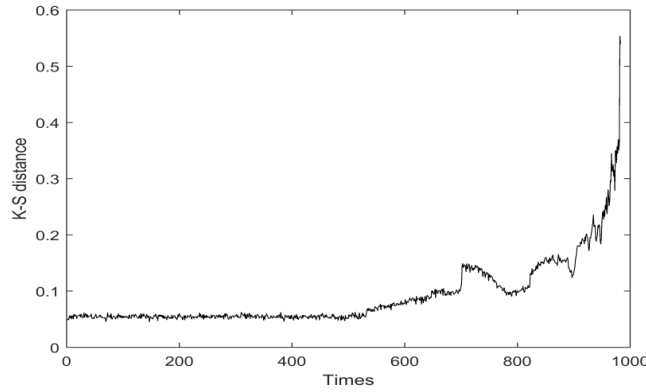


Figure 5. K-S distance of bearing vibration signal

4.2. Prediction of Bearing Remaining Useful Life

In this paper, eight K-S distance indexes are used to form the degradation sequence, namely, $N = 8$. Assuming that the early performance degradation index threshold K-S distance is 0.06514 and the failure threshold is 0.2669, the actual steps is 429 from the early degenerate point (the 533th collection point) to the failure point (the 962nd collection point) in the whole life experiment. Using LSTM prediction, when 413 steps prediction has been carried out, the K-S distance index reaches 0.2669.

Therefore, the estimated remaining useful life can be calculated as Equation (10).

$$\widehat{RUL} = P_{predicted} \times T = 413 \times 10 = 4130min \quad (10)$$

The actual remaining useful life can be calculated as Equation (11).

$$RUL = P_{actual} \times T = 429 \times 10 = 4290min \quad (11)$$

The prediction error is expressed by Equation (12).

$$error = \frac{|\widehat{RUL} - RUL|}{RUL} \times 100\% = 3.73\% \quad (12)$$

In order to compare the prediction results, the quadratic curve fitting method and the Grey prediction model (GM) method are used to predict the remaining useful life of the bearing at the same time, as shown in Figure 6.

According to Equation (13), the absolute average error (e_{MAE}), root mean square error (e_{RMSE}), normalized mean square error (e_{NMAE}), and average relative error (e_{MAPE}) of the three prediction methods can be calculated, and the results are shown in Table 1.

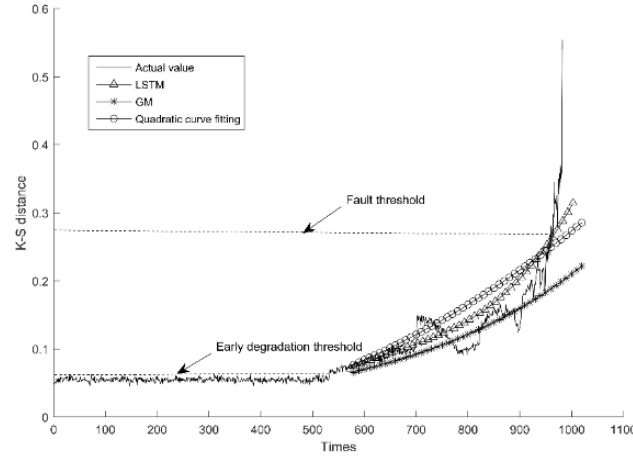


Figure 6. Prediction results of different methods

$$\left\{ \begin{array}{l} e_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ e_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ e_{NMAE} = \frac{1}{n\sigma^2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ e_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \end{array} \right. \quad (13)$$

Table 1. Error comparison of different prediction methods

Prediction methods	e_{MAE}	e_{RMSE}	e_{NMAE}	e_{MAPE}	Prediction steps	Life prediction error %
GM	0.0190	0.0241	0.0006	0.0282	555	29.37%
Quadratic curve fitting	0.0293	0.0370	0.0014	0.0946	458	6.76%
LSTM	0.0148	0.0206	0.0004	0.0277	413	3.73%

From Figure 6 and Table 1, LSTM has the best prediction effect and higher accuracy than the other two methods.

5. Conclusions

The objective of this work is to provide a simple and effective method for machinery RUL prediction under small sample conditions. At present, many online methods have been presented for machinery state diagnosis with high accuracy, which provide helpful technical support for timely machinery maintenance. However, from a management and economic benefit perspective, a group maintenance strategy [24-25] is usually employed in machinery maintenance, especially for batch machinery. Therefore, we demand not only the health state but also the RUL of machinery. In practice, in addition to small sample data, the full life data of machinery is often difficult to obtain. Under this condition, the method presented in this paper is applicable.

For future works, we plan to study the effect of sample size on RUL prediction accuracy. In addition, it would also be

interesting to work on the best maintenance strategy based on RUL prediction.

Acknowledgments

The original bearing full life vibration data was provided by the University of Cincinnati via its Intelligent Maintenance Center.

References

1. R. Rocchetta, M. Broggi, Q. Huchet, and E. Patelli, "Online Bayesian Model Updating for Structural Health Monitoring," *Mechanical Systems and Signal Processing*, Vol. 103, pp. 174-195, March 2018
2. G. Walter and S. D. Flapper, "Condition-based Maintenance for Complex Systems based on Current Component Status and Bayesian Updating of Component Reliability," *Reliability Engineering and System Safety*, Vol. 168, pp. 227-239, December 2017
3. J. H. Shin and H. B. Jun, "On Condition based Maintenance Policy," *Journal of Computational Design and Engineering*, Vol. 2, No. 2, pp. 119-127, January 2015
4. M. S. Kan, A. C. C. Tan, and J. Mathew, "A Review on Prognostic Techniques for Non-Stationary and Non-Linear Rotating Systems," *Mechanical Systems and Signal Processing*, Vol. 62-63, pp. 1-20, October 2015
5. J. L. Zhang, D. G. Shang, Y. J. Sun, and X. W. Wang, "Multiaxial High-Cycle Fatigue Life Prediction Model based on the Critical Plane Approach Considering Mean Stress Effects," *International Journal of Damage Mechanics*, Vol. 27, No. 1, pp. 32-46, January 2018
6. J. Q. Hu, L. B. Zhang, L. Ma, and L. Wang, "An Integrated Safety Prognosis Model for Complex System based on Dynamic Bayesian Network and Ant Colony Algorithm," *Expert Systems with Applications*, Vol. 38, No. 3, pp. 1431-1446, March 2011
7. L. Yang and J. Lee, "Bayesian Belief Network-based Approach for Diagnostics and Prognostics of Semiconductor Manufacturing System," *Robotics and Computer Integrated Manufacturing*, Vol. 28, No. 1, pp. 66-74, February 2012
8. Z. Q. Zhao, B. Liang, X. Q. Wang, and W. N. Lu, "Remaining Useful Life Prediction of Aircraft Engine based on Degradation Pattern Learning," *Reliability Engineering and System Safety*, Vol. 164, pp. 74-83, August 2017
9. M. A. Djeziri, S. Benmoussa, and R. Sanchez, "Hybrid Method for Remaining Useful Life Prediction in Wind Turbine Systems," *Renewable Energy*, Vol. 116, pp. 173-187, February 2017
10. J. Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic Modelling Options for Remaining Useful Life Estimation by Industry," *Mechanical Systems and Signal Processing*, Vol. 25, No. 5, pp. 1803-1836, July 2011
11. X. S. Si, W. Wang, C. H. Hu, and D. H. Zhou, "Remaining Useful Life Estimation—A Review on the Statistical Data Driven Approaches," *European Journal of Operational Research*, Vol. 213, No. 1, pp. 1-14, August 2011
12. L. X. Liao, "Discovering Prognostic Features using Genetic Programming in Remaining Useful Life Prediction," *IEEE Transactions on Industrial Electronics*, Vol. 61, No. 5, pp. 2464-2472, May 2014
13. J. B. Yu, "A Hybrid Feature Selection Scheme and Self-Organizing Map Model for Machine Health Assessment," *Applied Soft Computing*, Vol. 11, No. 5, pp. 4041-4054, July 2011
14. F. Liu, Y. B. Liu, F. L. Chen, and B. He, "Residual Life Prediction for Ball Bearings based on Joint Approximate Diagonalization of Eigen Matrices and Extreme Learning Machine," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, Vol. 231, No. 9, pp. 1699-1711, May 2015
15. W. Caesarendra, A. Widodo, and B. Yang, "Combination of Probability Approach and Support Vector Machine Towards Machine Health Prognostics," *Probabilistic Engineering Mechanics*, Vol. 26, No. 2, pp. 165-173, April 2011
16. C. Sun, Z. Zhang, and Z. He, "Research on Bearing Life Prediction based on Support Vector Machine and Its Application," *Journal of Physics: Conference Series*, Vol. 305, No. 1, 2011
17. S. J. Dong and T. H. Luo, "Bearing Degradation Process Prediction based on The PCA and Optimized LS-SVM Model," *Measurement*, Vol. 46, No. 9, pp. 3143-3152, November 2013
18. A. M. Fu, X. Z. Wang, Y. L. He, and L. S. Wang, "A Study on Residence Error of Training an Extreme Learning Machine and its Application to Evolutionary Algorithms," *Neurocomputing*, Vol. 146, pp. 75-82, December 2014
19. Y. N. Qian and R. Q. Yan, "Remaining Useful Life Prediction of Rolling Bearings using an Enhanced Particle Filter," *IEEE Transactions on Instrumentation and Measurement*, Vol. 64, No. 10, pp. 2696-2707, October 2015
20. L. Ren, J. Cui, Y. Q. Sun, and X. J. Cheng, "Multi-Bearing Remaining Useful Life Collaborative Prediction: A Deep Learning Approach," *Journal of Manufacturing Systems*, Vol. 43, pp. 248-256, April 2017
21. R. Pascanu, T. Mikolov, and Y. Bengio, "On the Difficulty of Training Recurrent Neural Networks," in *Proceedings of the 30th International Conference on Machine Learning*, No. part 3, pp. 1310-1318, June 2013
22. Y. Bengio, P. Simard, and P. Frasconi, "Learning Long-Term Dependencies with Gradient Descent Is Difficult," *IEEE Transactions on Neural Networks*, Vol. 5, No. 2, pp. 157-166, March 1994
23. "National Aeronautics and Space Administration," (<http://ti.arc.nasa.gov/c/3/>, accessed November 13, 2018)
24. N. Chalabi, M. Dahane, B. Beldjilali, and A. Neki, "Optimization of Preventive Maintenance Grouping Strategy for Multi-Component Series Systems: Particle Swarm Based Approach," *Computers & Industrial Engineering*, Vol. 102, pp. 440-451, December 2016
25. H. C. Vu, P. Do, and A. Barros, "A Stationary Grouping Maintenance Strategy using Mean Residual Life and the Birnbaum Importance Measure for Complex Structures," *IEEE Transactions on Reliability*, Vol. 65, No. 1, pp. 217-234, March 2016