ISSN: 2688-8645

Residual Life Prediction of Rolling Bearing based on AOA-LSTM

Tong Wang, Tao Xi*

School of Mechanical Engineering, Tiangong University, Tianjin 300387, China

Abstract

Aiming at the problem of predicting the remaining life of rolling bearings, a method for predicting the remaining life of bearings based on the arithmetic optimization algorithm (AOA) and the long-short-term memory network (LSTM) fusion algorithm is proposed. First, use the random forest algorithm to analyze the importance of the extracted time-domain and frequency-domain feature indicators, and build a degradation feature decision table; then, use the AOA optimization algorithm to optimize the hyperparameters in the LSTM, and select the optimal hyperparameters to establish predictions model; finally, the degradation features are input into the prediction model for prediction, and the prediction model is evaluated by root mean square error (RMSE) and mean absolute error (MAE). The proposed research method is experimentally verified by the XJTU-SY dataset. The experimental results show that the RMSE and MAE of the proposed research method are 5.56% and 4.37%, respectively. Compared with the MLP model, the RMSE and MAE are reduced by 31.58% and 29.61%, respectively. Compared with the RNN model, the RMSE and MAE are reduced by 24.66% and 25.49%, respectively, which verifies the effectiveness of the research method.

Keywords

Rolling Bearing; Remaining Life Prediction; Arithmetic Optimization Algorithms; Long Short-Term Memory Network.

1. Introduction

As a key component of mechanical equipment, rolling bearings will have a serious impact once their working conditions fail. The main reason for the failure of rolling bearings is due to the failure of bearings and the degradation of life during operation. In order to judge the health status of bearings before they reach the failure point, the degradation state assessment and remaining life prediction of bearings play an important role[1,2,3].

In recent years, the remaining useful life prediction of rolling bearings has become one of the hot research topics in the field of mechanical equipment fault diagnosis and health management. At present, the research field mainly uses data-driven methods to establish prediction models[4]. Xu et al. used the extracted high-dimensional degradation features into normalized indicators by PCA method, and input them into the regression SVR prediction model for bearing life prediction[5]. Hu used RVM to establish a prediction model to predict the remaining life of rolling bearings and achieved good prediction results. With the development of deep learning, the prediction model established by neural network has been widely used in the prediction of bearing life[6]. Wang proposed the life prediction method of long short-term memory network[7]. Yin used Bi-RNN method to predict the residual life of main bearing of wind turbine, and achieved high accuracy[8]. Han proposed a residual life prediction method for rolling bearings based on BiLSTM[9]. The above research is to predict the remaining life of the bearing by establishing a deep learning model, and has achieved good prediction results. However, in the process of model establishment, the parameters of the neural network model were selected through experience, there is a certain error between the selected model parameters and the optimal network parameters, which leads to a large error between the predicted remaining life in the model and the actual degradation life.

Therefore, a method for predicting the residual life of rolling bearings based on AOA-LSTM is proposed in this paper. Firstly, the time domain and frequency domain characteristic indexes are extracted from the original vibration signals. Then random forest algorithm was used to analyze the importance of the extracted high-dimensional feature indicators, and the feature indexes with high importance were selected to construct the degradation feature decision table. Finally, AOA optimization algorithm was used to optimize the superparameters in LSTM neural network, and the optimal superparameters were selected to establish a prediction model. The prediction model was evaluated by RMSE and MAE. The method is verified by XJTU-SY dataset, and good prediction results are obtained, which proves the feasibility and robustness of the proposed method.

2. Theoretical Basis

2.1. AOA Optimization Algorithm

Arithmetic optimization algorithm (AOA) is a new meta-heuristic intelligent optimization algorithm proposed by Abualigah in 2021[10]. The algorithm has good convergence speed and solution accuracy. The core principle is to use basic arithmetic operations to optimize the algorithm, where multiplication and division operations update the global optimal position to avoid local solutions. The addition and subtraction operation updates the local optimal position to improve the solution accuracy. The optimization process of arithmetic optimization algorithm is divided into three stages : initialization stage, exploration stage and development stage. Figure 1 shows the hierarchical structure of the algorithm. The outer layer is exploration stage, including division operation (D) and multiplication operation (M). The inner layer is the development stage, including addition operation (A) and subtraction operation (S), and the core point position is the optimal solution.

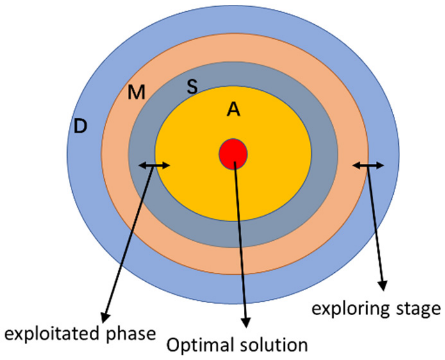


Figure 1. Hierarchical structure of AOA

(1) Initialization stage

Initialize parameters and randomly set the initial solution within the parameter range. The best candidate solution for each iteration is determined as the current optimal solution, which is :

$$X=\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n-1} & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n-1} & x_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N-1,1} & x_{N-1,2} & \cdots & x_{N-1,n-1} & x_{N,n-1} \\ x_{N,1} & x_{N,2} & \cdots & x_{N,n-1} & x_{N,n} \end{bmatrix} \tag{1}$$

Before the iteration of the algorithm, the exploration stage or the development stage is selected by the mathematical optimizer acceleration function (MOA). The random number $r_1 \in [0,1]$ is compared with the MOA, If $r_1 \geq MOA$ chooses the iterative optimization of the exploration stage, if $r_1 < MOA$ chooses the development stage, the MOA definition is:

$$MOA(\mathrm{t}) = Min + t \times (\frac{Max - Min}{M_t}) \tag{2}$$

Where, $MOA(t)$ is the function value of the *t-th* iteration, t is the current iteration number, M_t is the maximum iteration number, Max is the maximum function value of MOA, Min is the minimum function value of MOA.

(2) Exploration stage

In the exploration stage, two search mechanisms, division mechanism and multiplication mechanism, are mainly used as search strategies. The random number is compared with 0.5. If the outermost division strategy is selected, otherwise the multiplication strategy is selected. Equation (3) is the position update function. Equation (4) is the iterative process of the exploration stage :

$$MOP(t) = 1 - \frac{C_t^{\ 1/\alpha}}{M_t^{\ 1/\alpha}} \tag{3}$$

$$x_{i,j}(t+1) = \begin{cases} best(x_j) \div (MOP + \varepsilon) \times ((UB_j - LB_j) \times \mu + LB_j), & r2 < 0.5 \\ best(x_j) \times MOP \times ((UB_j - LB_j) \times \mu + LB_j), & r2 \geq 0.5 \end{cases} \tag{4}$$

Where, $MOP(t)$ is the function value at the *t-th* iteration, t is the current iteration number, M_t is the maximum iteration number, and α is the sensitive parameter, $x_i(t+1)$ represents the *i-th* solution of the next iteration, $x_{i,j}(t+1)$ represents the *i-th* solution at the *j-th* position of the current iteration, $best(x_i)$ represents the *j-th* position of the optimal solution of the current iteration, UB_j and LB_j represent the upper and lower bounds of the *j-th* position, ε is an integer, and μ is the control parameter of the adjustment search process.

(3) Development stage

In the development stage, addition mechanism and subtraction mechanism are used as search strategies for local optimization. In this stage, the search scope is small and the accuracy is high. In this stage, the search range is small and the accuracy is high. When the random number $r_1<MOA$, select this stage and set the random number $r_3 \in [0,1]$. If $r_3 < 0.5$ chooses the subtraction operation strategy, otherwise, select the addition operation strategy, the position update function of this stage is Equation (5) :

$$x_{i,j}(t+1) = \begin{cases} best(x_j) - MOP \times ((UB_j - LB_j) \times \mu + LB_j), & r3 < 0.5 \\ best(x_j) + MOP \times ((UB_j - LB_j) \times \mu + LB_j), & r3 \geq 0.5 \end{cases} \tag{5}$$

2.2. LSTM Algorithm and Parameter Optimization

Recurrent neural network (RNN) is suitable for processing time series data, but there are problems such as gradient disappearance and gradient explosion during model training. Long short-term memory network (LSTM) uses the principle of gating mechanism to make the model have stronger memory ability and can process longer time series data[11]. At the same time, it solves the gradient problem existing in RNN. The three gating mechanisms of LSTM are forgetting gate, input gate and output gate. The gating mechanism can compress the input data into a fixed interval and control the flow of data.

The structure of LSTM model is shown in Figure 2. The arrow direction is the direction of data flow; c_{t-1} and c_t are the memory data input of the last timestamp and output of the current timestamp respectively; h_{t-1} and h_t are the input data of the last timestamp and output data of the current timestamp respectively; x_t is the input data of the current timestamp.

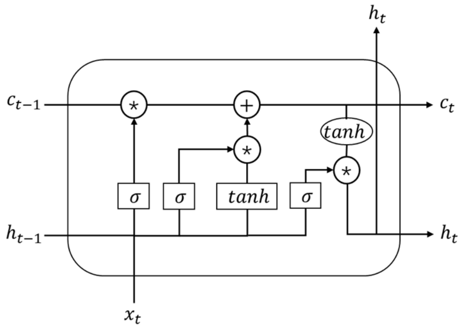


Figure 2. LSTM model structure diagram

The forgetting gate is the product of the input c_{t-1} of the previous timestamp and the forgetting gate control variable g_f of the current timestamp, and the output is $c_{t-1}g_f$.

$$g_f = \sigma(W_f \bullet [h_{t-1}, x_t] + b_f) \tag{6}$$

Where, W_f and b_f are the weight parameters and bias parameters of the forgetting gate, x_t is the input parameters of the current timestamp, and h_{t-1} is the output data of the timestamp.

The input gate is the product of the input gating value variable g_i and the state vector $\tilde{c}_t$ transformed by the tanh activation function of the current input data, and the output data is $g_i\tilde{c}_t$.

$$\tilde{c}_t = \tanh(W_c \bullet [h_{t-1}, x_t] + b_c) \tag{7}$$

Where, W_c is the weight parameter of the input gate, b_c is the bias parameter, and tanh is the activation function.

$$g_i = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i) \tag{8}$$

Where, W_i is the weight parameter of the input gate, and b_i is the bias parameter.

The output gate is the product of the output gating variable g_o and the output state vector of the current timestamp, and the output data is h_t.

$$g_o = \sigma(W_o [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = g_o \cdot \tanh(g_f c_{t-1} + g_i \tilde{c}_t) \tag{10}$$

Where, W_o is the weight parameter of the output gate, and b_o is the bias parameter.

The selection of parameters in the process of LSTM network training has a great impact on the results of network training. Most studies select LSTM network parameters through experience, but the obtained results will produce large errors. Therefore, this paper optimizes the hyperparameters in AOA optimization algorithm LSTM, and uses the optimization results to build the LSTM network prediction model. The important parameters in LSTM network include learning rate, number of iterations and number of neurons in each layer network, so AOA is used to find the global optimal value for the above parameters, and the Root Mean Square (RMSE) is selected as the fitness function.

$$RMSE(p, y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2} \tag{11}$$

Where, p is the predicted value, y is the true value, n is the number of predicted sample points.

Algorithm optimization steps are as follows:

(1) Initialize the parameters of AOA algorithm, set the maximum number of iteration, randomly select optimization parameters, and set the root-mean-square error as fitness function.

(2) Initialize the parameters of the LSTM algorithm, set parameters such as batch processing, and perform model prediction by initializing the optimization parameters to obtain the current iteration fitness value.

(3) AOA iteratively updates the parameter position, records each optimal result, and returns step (2).

(4) Calculate the optimal fitness value iteratively through steps (2) and (3), compare with the recorded value to determine whether the current optimal value is the optimal solution. If the iteration condition is met, the iteration is terminated, the optimal result is output. Otherwise, the update position is returned to steps (2) and (3) until the iteration condition is met.

2.3. Importance Analysis of Random Forest

The random forest algorithm (RF) is an ensemble learning algorithm based on decision tree[12]. The Bagging algorithm is used to randomly select new samples and features from the original data samples and features to form a sub-training set, and then the sub-training set is input into the model. Through multiple decision tree voting, the final classification or prediction results are obtained, and the importance score of each feature is obtained. Therefore, The RF algorithm is also used as a dimensionality reduction method for high-dimensional feature reduction. The main idea of random forest algorithm for feature importance analysis is to use the Out-Of-Bag data error (OOB) and the importance index of sample feature information. The out-of-bag error is calculated as Equation (12)

$$E_{OOB}=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}(y_r-y_p)^2 \tag{12}$$

Where, n is the total number of samples, y_r is the actual value of samples, and y_p is the calculated value of samples.

Assuming that the data sample for extracting multi-dimensional degradation features is N, and the accuracy of OOB on the i tree is A_i, the importance analysis of feature U is carried out, the value of OOB is calculated twice, and the importance value PI_i of feature U on the i tree is obtained, that is

$$PI_i(U)=E_{OOBi}-E'_{OOBi} \tag{13}$$

Then calculate the importance value of the feature corresponding to each decision number and calculate the mean value of the importance value to obtain the importance value of feature U.

$$PI(U)=\frac{1}{k}\sum_{i=1}^{k}PI_i(U) \tag{14}$$

Where, k is the number of decision trees.

2.4. Prediction Model Construction and Evaluation

The construction of AOA-LSTM prediction model is mainly divided into three stages, namely extracting features from original vibration signals and constructing degradation features, AOA optimization of LSTM hyperparameters, and establishment and evaluation of LSTM prediction model.

The construction of degradation features needs to extract the time domain and frequency domain feature indexes from the original data. Extraction of mean value, standard deviation, bias angle, kurtosis, maximum value, minimum value, peak-to-peak value, root mean square value, waveform factor, margin factor. Frequency domain degradation feature extraction center of gravity frequency, mean square frequency, frequency variance, energy a total of 14 dimensional degradation characteristics, feature extraction detailed formula reference [3]. Random forest algorithm was used to analyze the feature importance of the extracted high-dimensional feature index, and the features with higher correlation with the actual bearing degradation characteristics were selected to construct the degradation feature decision table, which was input into the LSTM prediction model for prediction and analysis.

In the construction of LSTM prediction model, the degradation feature decision table was used as the input data of the training set, and the corresponding life label was used as the training set label. In order to eliminate the influence of different dimensions, the degradation features were normalized. The remaining life label was initially 1, and as the bearing degradation process progresses, the residual life is finally 0. After the model training, the data of the test set were input, and then the model was evaluated by root Mean Absolute Error and mean absolute error (MAE).Average absolute error formula

$$MAE(p,y)=\frac{1}{n}\sum_{i=1}^{n}\left|p_i-y_i\right| \tag{15}$$

Where, *p* is the predicted value, *y* is the true value, *n* is the number of sample points.

3. Example Analysis and Verification

Firstly, the XJTU-SY dataset[13] is selected to verify the feasibility of the method, which includes 15 groups of full life cycle data of rolling bearings under three working conditions. Figure 3 shows the bearing acceleration life test platform. The bearing model selected in the experiment is LSD UER204 rolling bearing, with a sampling frequency of 25.6KHz and a sampling interval of 1min. The sampling time was 1.28s, that is, the sample point was 32768 per sampling. In the experiment, the load direction of the bearing is horizontal, and the acceleration vibration signal in the horizontal direction contains more effective degradation information. so the data in this direction is selected for test analysis.

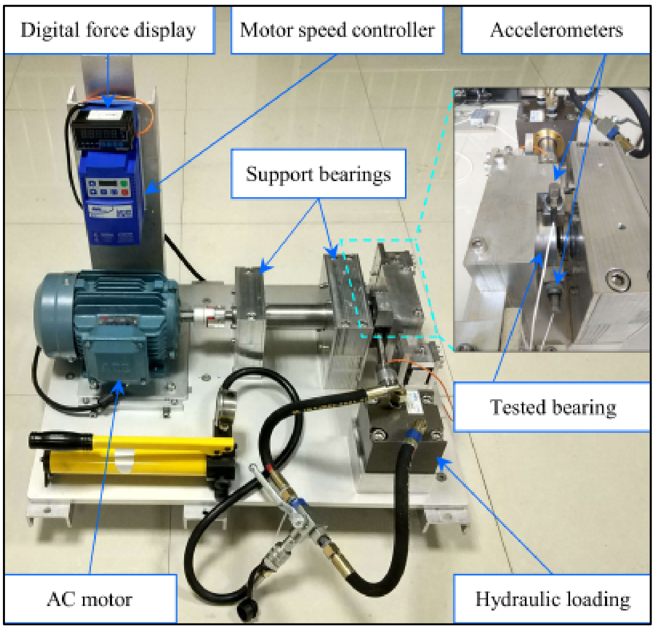


Figure 3. XJTU-SY bearing acceleration life test platform

Figure 4 is the time domain diagram and spectrum diagram of the full life bearing data of Bearing1-1. From Figure 4 (a), it is found that the whole life cycle process of the bearing is roughly divided into three degradation stages : normal wear stage, degradation stage and failure stage. Figure 4 (b) shows the spectrum of Bearing1-1. It is found that the spectrum energy of bearing degradation process is mainly concentrated in the 0-2000 Hz frequency band.

Bearing1-1 is selected to extract 14-dimensional feature indexes in time domain and frequency domain, and then the importance of the extracted feature indexes is analyzed by random forest algorithm. Figure 5 shows the relative significance analysis results. Table 1 shows the quantitative importance table of degradation features, A total of 5-dimensional features with importance values greater than 0.04 are selected to construct a degradation feature decision table as the input data of the LSTM model.

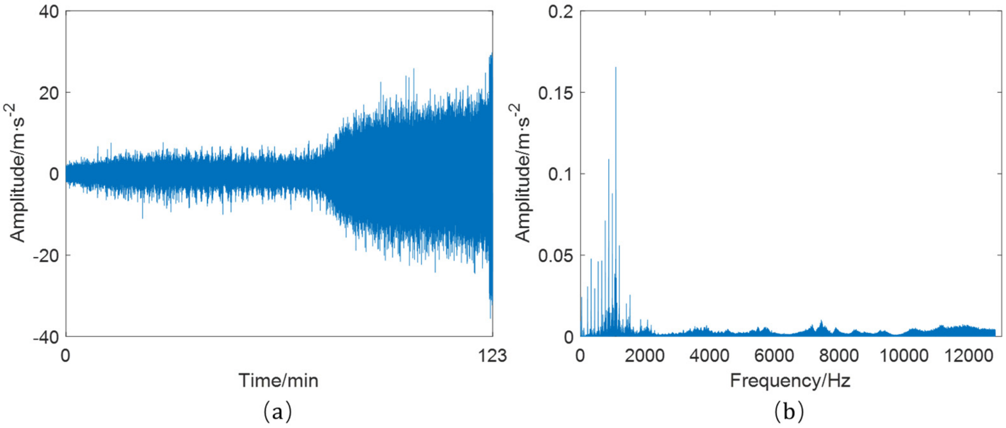


Figure 4. Bearing 1-1 time domain diagram and spectrum diagram

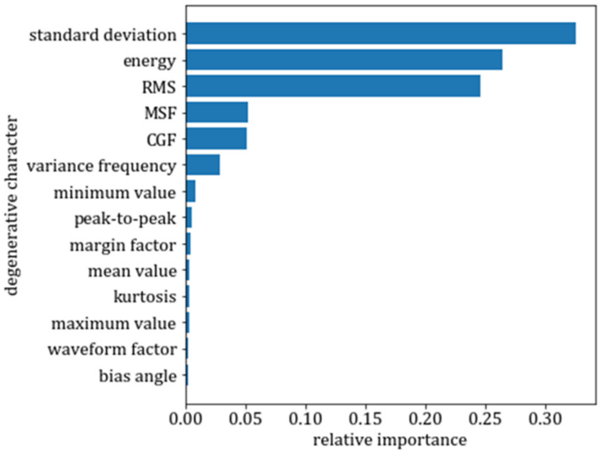


Figure 5. Relative importance of degradation features

Table 1. Relative importance of degradation features

characteristics	importance	Characteristics	importance
mean value	0.0038	RMS	0.2454
standard deviation	0.3246	waveform factor	0.0028
bias angle	0.0026	margin factor	0.0048
kurtosis	0.0031	center of gravity frequency	0.0509
maximum value	0.0029	mean square frequency	0.0520
minimum value	0.0088	variance frequency	0.0293
peak-to-peak	0.0051	energy	0.2639

The AOA optimization algorithm optimizes the number of network layers, the number of neurons in each layer, and the learning rate in LSTM as hyperparameters. The root mean square value is selected as the fitness function, and the AOA algorithm parameters are set to optimize the model hyperparameters. The optimization result is that the number of network layers is 2, the number of neurons in the first layer is 88, the number of neurons in the second layer is 35, and the learning rate is 0.0073.

The LSTM prediction model is constructed by the optimization results of the AOA optimization algorithm. Table 2 shows the parameter setting results of LSTM network model.

Table 2. LSTM network model parameter table

network parameter	value	network parameter	value
network layers	2	Input vector dimension	5
number of neurons	88, 35	activation function	Relu
learning rate	0.0073	optimizer	Adam
Batch size	128	Output vector dimension	1

Figure 6 shows the change process of MSE loss curve during model training. The number of iterations is set to 1000. It can be found from the figure that the training loss curve and the verification loss curve tend to be stable with the increase of the number of iterations.

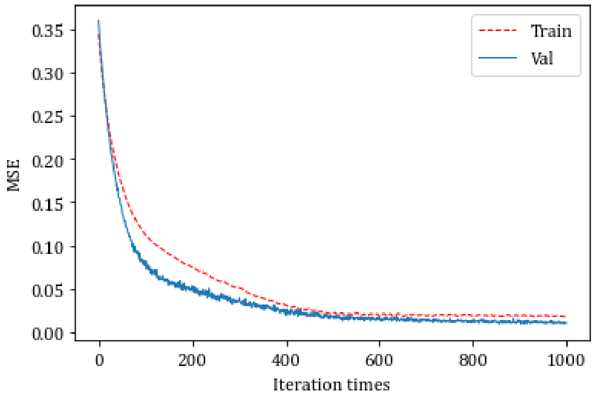
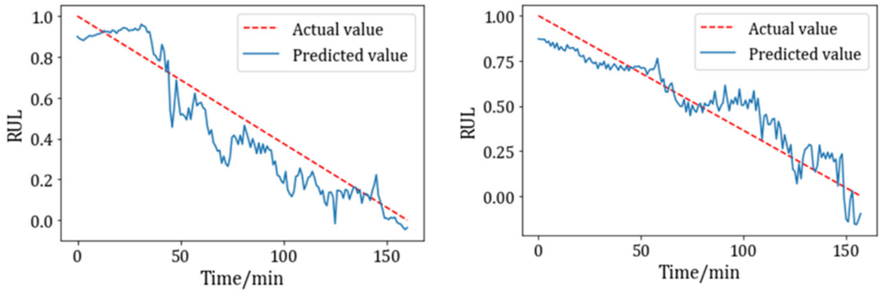
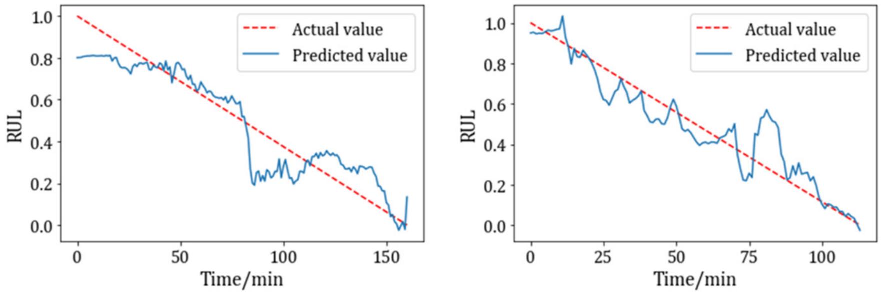


Figure 6. Loss curve change process diagram

Taking Bearing1-1 as the training set, the remaining four bearings are selected as the test set. Figure 7 shows the prediction results of bearings under different working conditions. The red curve is the actual life degradation process, and the blue curve is the predicted value. It is found from the diagram that the predicted value fluctuates with the actual value within a certain error range, and a better prediction result is obtained.



(a) Bearing1-2 (b) Bearing1-3



(c) Bearing2-2 (d) Bearing3-5

Figure 7. AOA-LSTM prediction results

Table 3 is the evaluation table of the prediction model. RMSE and MAE are used to evaluate the prediction model. The evaluation results are shown in Table 3.

Table 3. LSTM network model parameter table

data set	RMSE	MAE	data set	RMSE	MAE
Bearing1-2	0.1169	0.0965	Bearing2-2	0.1749	0.1484
Bearing1-3	0.1038	0.0875	Braring3-5	0.0930	0.0690

In order to verify the effectiveness of the proposed method, MLP, RNN algorithm and AOA-LSTM algorithm are selected for comparison. Figure 8 shows the comparison of the prediction results of three different algorithms. It can be concluded that the proposed method has a higher degree of fitting with the actual prediction results.

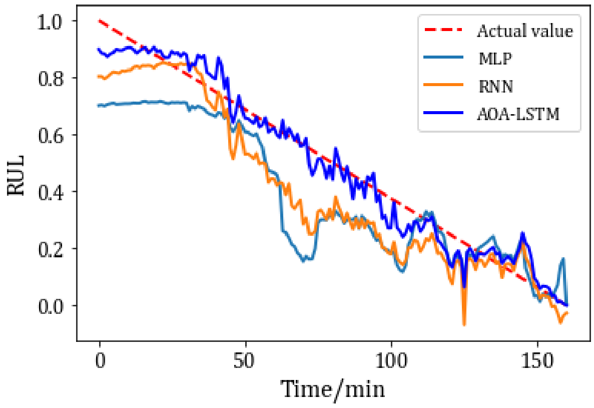


Figure 8. Comparison of prediction results of different prediction models

Table 4 uses RMSE and MAE to evaluate the errors of three different prediction models. It can be seen that the error of AOA-LSTM prediction model is smaller, which proves the effectiveness of the proposed method.

Table 4. Evaluation of different prediction models

model	RMSE	MAE
MLP	0.3714	0.3398
RNN	0.3022	0.2986
AOA-LSTM	0.0556	0.0437

4. Conclusion

In this paper, a residual life prediction method of rolling bearing based on AOA-LSTM is proposed. Through the verification of XJTU-SY data set, more accurate prediction results are obtained. Firstly, the random forest algorithm is used to analyze the importance of the extracted feature indexes, and then the AOA optimization algorithm is used to find the optimal value of the hyperparameters of LSTM, and the prediction model is established. The following conclusions are obtained:

(1) The time domain and frequency domain feature indexes are extracted from the original vibration signal, and the importance analysis is carried out by using the random forest algorithm. Therefore, the selected degradation characteristics are more correlated with the actual residual life degradation characteristics.

(2) The AOA optimization algorithm is used to find the optimal parameter combination of the LSTM neural network, and the rolling bearings under different working conditions in the XJTU-SY data set is predicted. The prediction results are more accurate. Meanwhile, compared with the MLP and RNN models, the prediction error of the proposed method is smaller.

(3) The RMSE and MAE of the prediction results on the XJTU-SY dataset are 5.56 % and 4.37 %, respectively. The prediction results have high accuracy, which proves the feasibility of the proposed method.

References

[1] D Wang, K L Tsui, Q Miao. Prognostics and health management: A review of vibration based bearing and gear health indicators[J]. IEEE Access, Vol. 6 (2017), p.665-676.

[2] Z Q Chen, X D Chen, José Valente de Olivira, et al. Application of deep learning in equipment prognostics and health management[J].Chinese Journal of Scientific Instrument, Vol. 40 (2019) No.9, p.206-226.

[3] L Xu, X T Zheng, B Fu, et al. Fault Diagnosis Method of Motor Bearing Based on Improved GAN Algorithm[J]. Journal of Northeastern University(Natural Science), Vol. 40 (2019) No.12, p.1679-1684.

[4] H Pei, C H Hu, X S Si, et al. Review of machine learning based remaining useful life prediction methods for equipment[J]. Journal of Mechanical Engineering, Vol. 55 (2019) No.8, p.1-13.

[5] Z C Xu, L J Wang, Y Liu, et al. A Prediction Method for Remaining Life of Rolling Bearing Using Improved Regression Support Vector Machine[J]. Journal of Xi'an jiaotong University, Vol. 56 (2022) No.3, p.197-205.

[6] X M Hu, Y Wang, Z C Ji, et al. Fuzzy Information Granulation and Improved RVM for Rolling Bearing Life Prediction[J]. Journal of System Simulation, Vol. 33 (2021) No.11, p.2561-2571.

[7] F Wang, X Liu, G Deng, et al. Remaining life prediction method for rolling bearing based on the long short-term memory network[J]. Neural processing letters, Vol. 50 (2019) No.3, p.2437-2454.

[8] S Yin, G L Hou, X D Yu, et al. Research on temperature early warning method of wind turbine main bearing based on Bi-RNN[J]. Journal of Zhengzhou University (Engineering Science),Vol. 40 (2019) No.5, p.45-51.

[9] L J Han, C P Shi, J C Zhang, et al. Prediction for remaining useful life of rolling bearings based on BiLSTM[J].Manufacturing Automation, Vol. 42 (2020) No.5, p.47-50.

[10] L Abualigah, A Diabat, S Mirjalili, et al. The arithmetic optimization algorithm[J]. Computer methods in applied mechanics and engineering, Vol. 376 (2021), p.113609.

[11] Y Yu, X Si, C Hu, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural computation, Vol. 31 (2019) No.7, p.1235-1270.

[12] Q Zhou, H Zhou, Q Zhou, et al. Structure damage detection based on random forest recursive feature elimination[J]. Mechanical Systems and Signal Processing, Vol.46 (2014) No.1, p.82-90.

[13] B Wang, Y Lei, N Li, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings[J]. IEEE Transactions on Reliability, Vol. 69 (2018) No.1, p.401-412.