Macroscopic-Microscopic Attention in LSTM Networks based on fusion Features for prediction of bearing remaining life

QIN Yi[1,2] XIANG SHENG[1,2] HAORAO YAN[1,2]

[1] College of Mechanical Engineering, Chongqing University, Chongqing 400044, China
qy_808@aliyun.com
a1121623518@1636.com
[2] State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, China
18334729055@163.com

Abstract

In the mechanical transmission system, the bearing is one of the most widely used transmission components. The failure of the bearing will cause serious accident and huge economic loss. Therefore, the remaining life prediction of the bearing is greatly important. In order to predict the remaining life of the bearing, a prediction method combining macro-micro attention, long-short-term memory neural network and isometric mapping is proposed. First, some typical time-domain and frequency-domain characteristics of the vibration signal are calculated .respectively, such as the the mean frequency, the absolute mean value, the standard deviation, the RMS and so on. Then, the principal component of these characteristics is extracted by the isometric mapping method. The importance of fusional characteristic information is filtered via a proposed macro-micro attention mechanism, so that the input weight of neural network data and recursive data can reach multi-level real-time amplification. With the new long short-term memory neural network, the characteristic of the bearing vibration signal can be predicted based on the known fusional characteristic. The experimental results show that the method can predict the remaining life of the bearing well and has higher prediction accuracy than the conventional LSTMs.

1 Introduction

Bearing is widely used in mechanical equipment and is one of the most universally used mechanical parts[1]. Under the complex working condition and environment, bearings are easily subject to failures, which may result in the catastrophe of the machine running and even threaten the personal safety [2]. Bearing remaining life prediction is beneficial to determine the equipment maintenance time reasonably, improve the production efficiency, reduce the accident rate, and prevent the sudden accidents, which is significant for engineering production [3].Recently, with the rapid developments in sensing, signal processing and artificial intelligence technology, Prognostics and Health Management (PHM) technologies based on data (experience) have gradually become the mainstream solution either in fault diagnosis or remaining useful life（RUL）estimation, instead of physics-based methods which can be expensive and tedious to develop[4]. Neural network is one of the most advanced models for the study of sequence classification and prediction in data driven methods. Zhao et al. [5] introduced the application of deep learning in machine health monitoring system, mainly from the perspective of automatic encoder (AE) and its variants, restricted boltzmann machine and its variants, including deep belief network (DBN) and deep boltzmann machine (DBM), convolutional neural network (CNN) and recursive neural network (RNN). Qin et al. [6] proposed a model for fault diagnosis of wind turbines' gearboxes based on deep belief networks (DBNs) with improved logistic sigmoid units via extracting impulsive features. Gebraeel et al. [7] used the amplitude of the bearing vibration decay signal as the criterion to assess the life, and used the BP neural network to predict the life.

Based on particular design, RNN is suitable for processing timing-related information [8]. However, RNN also has its own drawbacks, e.g. the excessive recursion time is indirectly equivalent to increasing the depth of neural network and the training time; the vanishing gradient problem usually [9]. In order to solve these problems, long short-term memory (LSTM) was proposed by Hochreiter and Schmidhuber in 1997 [10]. It can avoid the long-term dependence problem of RNN, thus it has been widely used. Yuan et al. [11] investigated three RNN models including vanilla RNN, LSTM and GRU models for fault diagnosis and prognostics of aero engine. They found that these advanced RNN model based on LSTM and GRU performed better than the convetional RNN via a number of experiments. Elsheikh et al. [12] proposed the bidirectional handshake network which solved the problem that bidirectional LSTM could not be well utilized in the prediction field, and proved the superiority of the method in the life experiment of turbine engines.

Attention mechanism can be regarded as a kind of contribution screening of information which improves the efficiency of neural network by selecting key information for processing. Attention-based recurrent networks have been successfully applied to a wide variety of tasks, such as handwriting synthesis[13], machine translation[14], image caption generation[15] and visual object classification[16]. In the prediction field, attention-based LSTM is getting more and more attention. Ran et al. [17] substituted a tree structure with attention mechanism for the unfolding way of standard LSTM to construct the depth of LSTM and modeling long-term dependence for travel time prediction. Fernando et al. [18] combined two kinds of attention and used LSTM for human trajectory prediction and abnormal event detection. Filtering key information can reduce computing resources, but it can also cause some degree of information loss. Differential treatment of input data according to the screening of attention mechanism can not only reflect the focus to important information, but also retain useful information as far as possible.The importance of fusional characteristic information is filtered via a proposed macro-micro attention mechanism, so that the input weight of neural network data and recursive data can reach multi-level real-time amplification. With the new long short-term memory neural network, the characteristic of the bearing vibration signal can be predicted based on the known fusional characteristic. The experimental results show that the method can predict the remaining life of the bearing well and has higher prediction accuracy than the conventional LSTMs.

2 Related work

2.1 LSTM

The LSTM neural network is a special RNN neural network which also consists of an input layer, a hidden layer, and an output layer. The difference lies in using an LSTM structure that includes the input gate, the output gate, the forget gate, and the memory cell as the hidden layer, as shown in Figure. 1.

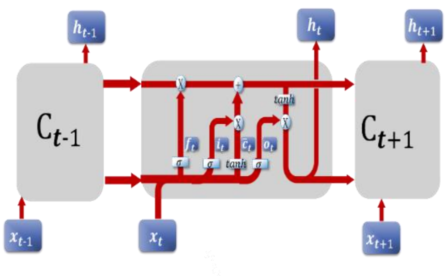


Figure. 1 Hidden layer structure of LSTM neural network

The forget gate is used to determine whether to keep the historical information stored in the current memory cell. If the door is opened, the historical information stored in the current memory cell is retained, otherwise the historical information is forgotten. The input gate is used to determine whether to allow the input layer information to enter the current memory cell. The open door allows the input layer signal to enter, and the closed door does not allow. The output gate is used to determine whether to output the current input layer signal to the next layer, the open door allows signal output and the closed door does not allow.

An LSTM network computes a mapping from an input sequence $\mathrm{x} = (x_1, x_2, \ldots x_n)$ to an output sequence $\mathrm{y} = (y_1, y_2, \ldots y_n)$ by calculating the network unit activations using the following equations iteratively from t = 1 to T:

$$
\begin{aligned}
\mathrm{i}_t &= \sigma\left(\mathrm{w}_{ix}\mathrm{x}_t + \mathrm{w}_{im}\mathrm{h}_{t-1} + \mathrm{b}_i\right) \\
\mathrm{f}_t &= \sigma\left(\mathrm{w}_{fx}\mathrm{x}_t + \mathrm{w}_{fm}\mathrm{h}_{t-1} + \mathrm{b}_f\right) \\
\mathrm{c}_t &= \mathrm{f}_t \quad \mathrm{c}_{t-1} + \mathrm{i}_t \quad \tanh\left(\mathrm{w}_{cx}\mathrm{x}_t + \mathrm{w}_{cm}\mathrm{h}_{t-1} + \mathrm{b}_c\right) \\
\mathrm{o}_t &= \sigma\left(\mathrm{w}_{ox}\mathrm{x}_t + \mathrm{w}_{om}\mathrm{h}_{t-1} + \mathrm{b}_o\right) \\
\mathrm{h}_t &= \mathrm{o}_t \quad \tanh\left(\mathrm{c}_t\right)
\end{aligned} \tag{1}
$$

where, i is the input gate, o is the output gate, f is the forgetting gate, c includes cell activation vectors, and h is the memory cell outputs. w represents the weight matrix (for example, the weight w_{ix} matrix representing the input x to the input gate), and b represents the threshold (b_i which is the threshold of the input gate). σ is the sigmoid activation function, $\tanh$ is the tanh activation function, represents dot product.

3 Macroscopic-Microscopic Attention in LSTM

The structure of MMALSTM neural network is in Figure. 2. The number of input cells and the number of output cells in MMALSTM are respectively set as 60 and 1, and the learning rate is set as 0.05. The number of hidden layer cells is set as 17 in this study. The initialization method of neural network employs the standard initialization. And the recurrent neural network based on MMA is deduced as follows.

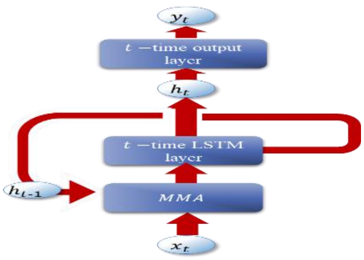


Figure. 2 structure of MMALSTM neural network

Firstly, we deal with the data matrix and calculate its macro and micro attention coefficient by the macro -micro attention mechanism. The input information $\mathbf{X}_t = \left[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_t\right]$ at time t is defined as $\mathbf{x}_t = \left[x_{t,1} \quad x_{t,2} \quad \dots \quad x_{t,n}\right]$ and the recurrent information at time t-1 is defined as $\mathbf{h}_{t-1} = \left[h_{t-1,1} \quad h_{t-1,2} \quad \dots \quad h_{t-1,m}\right]$. Macro attention mechanism is to process the data in the whole time interval by attention mechanism. Micro attention mechanism is to process the data $\mathbf{x}_t = \left[x_{t,1} \quad x_{t,2} \quad \dots \quad x_{t,n}\right]$ and $\mathbf{h}_{t-1} = \left[h_{t-1,1} \quad h_{t-1,2} \quad \dots \quad h_{t-1,m}\right]$ in each time instant by attention mechanism. Thus the operation of attention mechanism on data in the whole time dimension and each time dimension is called macro-micro attention mechanism. In this paper, the macro -micro attention mechanism of input matrix and the micro attention mechanism of recurrent matrix are processed (The recurrent data in the whole time dimension is not known before input into the network, so it cannot be processed at the macro level). In the training process, $\overline{x}_{T+1}$ is set as query vector q_M at the macro level, $x_{t+1.n}$ is set as query vector q_m at the micro level. In the prediction process, $\overline{x}_T$, $x_{t.n}$ represents q_M and q_m separately. And then the attention coefficients are calculated by:

$$
\chi_i = \frac{\exp\left(S\left(\overline{x}_i, \mathrm{q} \;\right)\right)}{\sum_{j=1}^{T} \exp\left(S\left(\overline{x}_j, \mathrm{q} \;\right)\right)} \tag{2}
$$

$$
\alpha_i = \frac{\exp\left(s\left(x_{t,i}, \mathrm{q}\right)\right)}{\sum_{j=1}^{n} \exp\left(s\left(x_{t,j}, \mathrm{q}\right)\right) + \sum_{p=1}^{m} \exp\left(s\left(h_{t-1,p}, \mathrm{q}\right)\right)} \tag{3}
$$

$$
\lambda_i = \frac{\exp\left(s\left(h_{t,i}, \mathrm{q}\right)\right)}{\sum_{j=1}^{n} \exp\left(s\left(x_{t,j}, \mathrm{q}\right)\right) + \sum_{p=1}^{m} \exp\left(s\left(h_{t-1,p}, \mathrm{q}\right)\right)} \tag{4}
$$

where α_i , λ_i are micro attention coefficients corresponding to input data and recurrent data respectively, χ_i is macro attention coefficients corresponding to the whole input data, $\bar{x}_i$ represents the mean value of $\mathbf{x}_i$ .And

$$S(\bar{x}_i, q) = \frac{\bar{x}_i q}{\sqrt{t}} \tag{5}$$

$$s(x_i, q) = \frac{x_{t,i}^T q}{\sqrt{n+m}} \tag{6}$$

$$s(h_i, q) = \frac{h_{t-1,i}^T q}{\sqrt{n+m}} \tag{7}$$

Secondly, according to the macro -micro attention coefficient and Eq. (4), the weight of input data is amplified in real time at multiple levels. And the weight of recursive data is amplified in real time based on the micro attention coefficient.

$$\begin{aligned}
\mathbf{w}'_{ix} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{ix} \\
\mathbf{w}'_{ox} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{ox} \\
\mathbf{w}'_{fx} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{fx} \\
\mathbf{w}'_{ih} &= (1+\lambda_i)\mathbf{w}_{ih} \\
\mathbf{w}'_{oh} &= (1+\lambda_i)\mathbf{w}_{oh} \\
\mathbf{w}'_{fh} &= (1+\lambda_i)\mathbf{w}_{fh}
\end{aligned} \tag{8}$$

From the above, the flow chart of MMA, MA, ma are depicted in the follow figures. Figure. 3 represents the flowchart of MMA, Figure. 4 (a)represents the flowchart of MA, Figure. 5(b) represents the flowchart of ma.

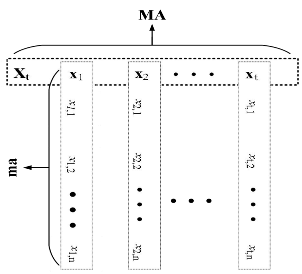
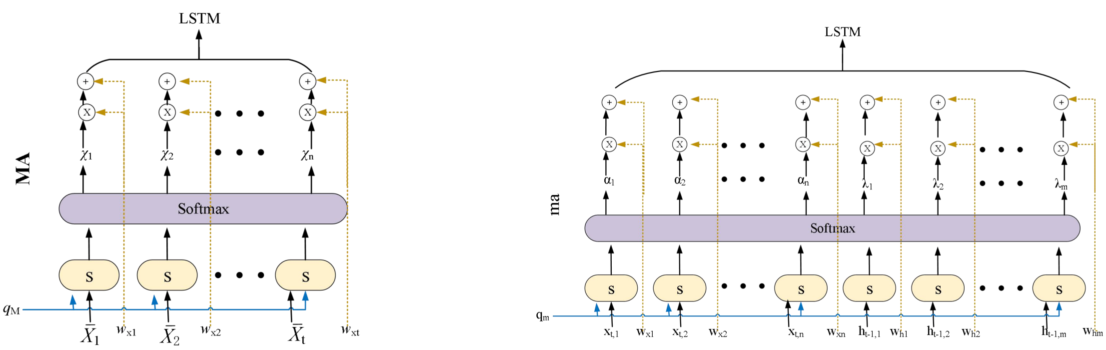


Figure. 3 The flowchart of MMA



(a) (b)

Figure. 4 The flowchart of MA and ma.

With the weight amplified based on MMA, a new variant of LSTM is proposed, which is named as MMALSTM. Via Eq. (1), we can derive the calculation formula of MMALSTM as follows:

$$
\begin{aligned}
&\mathbf{i}_t = \sigma\left(\mathbf{w}'_{ix}\mathbf{x}_t + \mathbf{w}'_{ih}h_{t-1} + \mathbf{b}_i\right) \\
&\mathbf{f}_t = \sigma\left(\mathbf{w}'_{fx}\mathbf{x}_t + \mathbf{w}'_{fh}h_{t-1} + \mathbf{b}_f\right) \\
&\mathbf{c}_t = \mathbf{f}_t \quad \mathbf{c}_{t-1} + \mathbf{i}_t \quad \tanh\left(\mathbf{w}_{cx}\mathbf{x}_t + \mathbf{w}_{ch}\mathbf{r}_{t-1} + \mathbf{b}_c\right) \\
&\mathbf{o}_t = \sigma\left(\mathbf{w}'_{ox}\mathbf{x}_t + \mathbf{w}'_{oh}h_{t-1} + \mathbf{b}_o\right) \\
&\mathbf{h}_t = \mathbf{o}_t \quad \tanh\left(\mathbf{c}_t\right) \\
&\mathbf{y}_t = \mathrm{g}\left(\mathbf{w}_{yh}h_t + \mathbf{b}_y\right)
\end{aligned}
\tag{9}
$$

where g is the linear activation function.

4 Prediction method based on MMALSTM

In summary, the steps of the proposed method for predicting the bearing remaining life based on the combination of the isometric mapping algorithm [19] and MMALSTM neural network are given below:

1. During the bearing life cycle, collect the bearing vibration samples according to the interval Ts between two adjacent samples. Suppose that the number of sample is n .

2. Calculate the 5 kinds of time-frequency characteristics （time-domain :mean absolute difference，standard deviation and root-mean-square, frequency-domain: mean of frequency distribution, envelope spectrum: mean of frequency distribution）of these vibration samples separately, then the eigenvalue matrix X of the size $n \times 5$ can be obtained.

3. Select the eigenvalue matrix $X1$ of the $n1$ samples from X as the training matrix.

4. The training matrix $X1$ and original matrix X are respectively processed by the ISOMAP algorithm and Savitzky-Golay filtering, and the obtained eigenvectors with the largest eigenvalues after filtering $V1 = \left(v1_1, v1_2, \dots, v1_{n1}\right)^T$ and $V = \left(v_1, v_2, \dots, v_n\right)^T$ are used as their principal components respectively. The matrix X and the vector V are only used to verify the validity of this method, which don't participate in neural network training and prediction.

5. Since the size of the matrix X is larger than that of the matrix $X1$ and both the sums of V and $V1$ are equal to zero according to the characteristic of ISOMAP algorithm, the two vectors may have different starting value, even though their trends are same. Therefore, it is necessary to unify them. By minimizing

$$
E = \sum_{i=1}^{n1}\left(v1_i - av_i - b\right)^2 \tag{10}
$$

a and b can be computed, and then all the elements of V are unified through the following equation

$$
v_i^{'} = av_i + b \tag{11}
$$

6. Linearly normalize the vector $V1$ to obtain the normalized vector $W = \left(w_1, w_2, \dots, w_{n1}\right)^T$.

7. Reconstruct matrix U :

$$
U = \begin{bmatrix} w_1 & w_2 & \cdots & w_{n1-p} \\ w_2 & w_3 & \cdots & w_{n1-p+1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p+1} & w_{p+2} & \cdots & w_{n1} \end{bmatrix} = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \vdots \\ \mathbf{u_{p+1}} \end{bmatrix} \tag{12}
$$

where p is the cell number of input layer and

$$
\mathbf{u}_i = \left[w_i \quad w_{i+1} \cdots \quad w_{n1-p+i-1}\right] \tag{13}
$$

8. Set the first p vectors of the matrix U as the input of the MMALSTM neural network and the last vectors as the output respectively, then train the MMALSTM neural network.

9. With the trained MMALSTM neural network, the mapping function f for prediction is defined. By inputting the last p vectors of the matrix U into the trained MMALSTM, the output $\overline{u}_{p+2}$ at the first prediction time (FPT) can be calculated as

$$
\overline{\mathbf{u}}_{p+2} = f\left(\mathbf{u}_2, \mathbf{u}_3 \cdots, \mathbf{u}_{p+1}\right) \tag{14}
$$

Then U is updated by

$$U=[\mathbf{u_1}\quad \mathbf{u_2}\quad \cdots\quad \mathbf{u_{p+1}}\quad \bar{\mathbf{u}}_{\mathbf{p+2}}]^T \tag{15}$$

10. Repeating step 9 several times, we can obtain the predicted data series $\bar{u}_i$.

$$\begin{aligned}
&\bar{\mathbf{u}}_{p+3}=f\left(\mathbf{u}_3,\mathbf{u}_4\cdots,\mathbf{u}_{p+1},\bar{\mathbf{u}}_{p+2}\right)\\
&\bar{\mathbf{u}}_{p+4}=f\left(\mathbf{u}_4,\mathbf{u}_5\cdots,\mathbf{u}_{p+1},\bar{\mathbf{u}}_{p+2},\bar{\mathbf{u}}_{p+3}\right)\\
&\bar{\mathbf{u}}_{p+5}=f\left(\mathbf{u}_5,\mathbf{u}_6\cdots,\mathbf{u}_{p+1},\bar{\mathbf{u}}_{p+2},\bar{\mathbf{u}}_{p+3},\bar{\mathbf{u}}_{p+4}\right)\\
&\qquad\vdots\\
&\bar{\mathbf{u}}_i=f\left(\mathbf{u}_{i-p},\mathbf{u}_{i-p+1}\cdots,\mathbf{u}_{p+1},\bar{\mathbf{u}}_{p+2}\cdots\bar{\mathbf{u}}_{p+(i-p-2)},\bar{\mathbf{u}}_{p+(i-p-1)}\right)
\end{aligned} \tag{16}$$

When the predicted characteristic data $\bar{w}_{n1-p+i-1}$ ($\bar{w}_{n1-p+i-1}\in\bar{\mathbf{u}}_i$) exceeds a preset threshold, the bearing remaining life can be calculated by $(i-n1)\times T_s$. Then we anti-normalize the predicted data series and get $\bar{v}_i$ ($i=n1+1,n1+2,\cdots,n$), which can be compared with the actual vector $V'=\left(v'_{n1+1},v'_{p+2},\ldots,v'_n\right)^T$ to verify the validity of this method.

5 Experimental signal analysis

The experimental data come from PRONOSTIA in the IEEE PHM 2012 Data Challenge [20]. The platform mainly contains three major parts: a rotatory part, a degradation generation part, and a signal acquisition part. To accelerate the degradation of bearing, radial load force is applied with a controllable shaft speed. Two accelerated sensors perpendicular to each other are installed on the key position for the test bearing. The sampling frequency is 25 600 Hz. Each sample has a duration of 0.1 s, which means each sample has 2560 points. The record interval is 10 s. The test is ceased once the amplitude of the collected signal surpasses a certain level to prevent damage. Our proposed method is applied on the data set bearing1-1, which means that our method currently only considers the operation condition with constant speed and load. The run-to-failure bearing1-1 contains 2803 samples. Moreover, many samples were acquired under the steady stage, and their characteristics were almost the same, which were of little significance to use these data for prediction. Therefore, only the last 1750 samples were used for bearing life prediction.

The characteristic matrix of 1650 experimental samples is used as the training matrix to predict the 100 characteristic points corresponding to the next 100 time instants. Via the proposed method, the training, predictive and actual curves are illustrated in Figure. 7. As can be seen from Figure. 5, the trend of the predicted curve is not close to actual curves but both have down trend. It follows that the predicted value only exceed the threshold when the MMALSTM neural network predict a large number of points. In such case, the predicted sampling points are 1736, which is 14×10s different from the actual life. In this study, to assess the performance of the proposed approach, an error index is defined as

$$Er=\frac{Rul-\overline{Rul}}{Rul}\times 100\% \tag{17}$$

where Rul denotes the actual remaining useful life, and Rul denotes the predictive remaining useful life. It is easy to note that the prediction error Er is 9.3%.

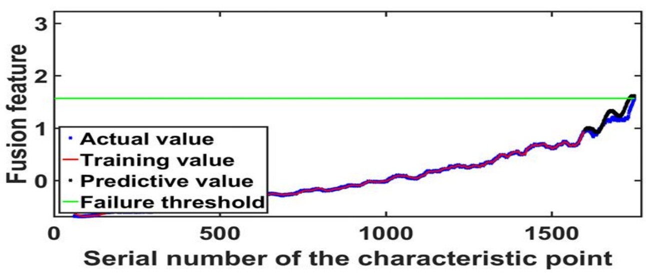


Figure. 5 Failure threshold, training curve, predicted curve and actual curve for 370 experimental samples of real bearing.

Then we add the number of known characteristic points to test the predictive power of the model, each group was tested for 5 times, and the mean percentage error of predicted RUL was calculated, as shown in the Figure .6.

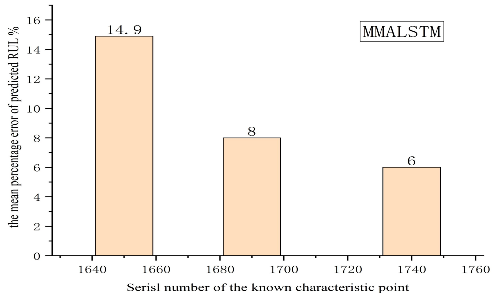


Figure. 6 The obtained prediction errors under the various numbers of the known characteristic point of data set .

Next, given 1690 characteristic points, MMALSTM, LSTM, deep LSTM (DLSTM) and LSTM with a projection (LSTMP) are used for comparison under the same conditions, each method was tested for 5 times, and the mean percentage error of predicted RUL was calculated, as shown in the Figure .7.

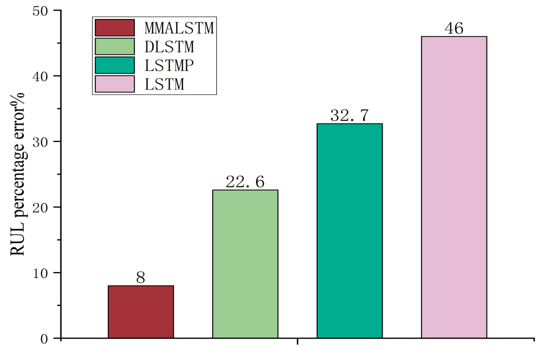


Figure. 10 Comparison of bearing remaining useful life prediction results obtained by the four types of LSTMs.

From the experimental results, it can be concluded that MMALSTM has not higher accuracy, while the four types of LSTMs do not have the satisfactory performance when fewer samples are used. And the closer the prediction point is to the failure point, the more accurate the prediction of bearing remaining life will be. Therefore, the proposed method is more suitable for bearing life prediction with enough samples.

6 Conclusion

Modern manufacturing requires high reliability and high efficiency, which makes bearing health analysis and residual life prediction an increasingly important research field. Advances in networked manufacturing and AI-oriented big data analysis provide new opportunities for bearing health analysis and residual life prediction. In this paper, isometric mapping algorithm is used to fuse 3 time-domain features,1 frequency-domain features and 1 envelope spectrum features of bearing vibration signal into a new feature. Due to the information characteristics of life data, the input weight of neural network and the weight of recurrent layer are all multi-amplified in real time by partial processing based on MMA. We use the improved long and short term neural network(MMALSTM) to study the bearing degradation model and predict the residual service life. The performance of the network model is verified by the data obtained from the bearing lifetime experiment and compared with LSTM, DLSTM and LSTMP. Numerical experiment results show that the proposed method not only has a higher accuracy of predicted. This will not only have certain reference significance and help for the online detection of bearing residual life, but also have great significance for the determination of equipment maintenance time and the prevention of unexpected accidents in engineering production.

In the future we will plan to study the performance of the proposed depth structure model in the field of bearing residual life prediction. Furthermore, the prediction of the residual life of bearings under variable working conditions by the deep learning model is also worth studying and working on

references

[1]Yi Qin, Jingqiang Zou, Baoping Tang, Yi Wang, Haizhou Chen. Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary, IEEE Transactions on Industrial Informatics, 2019, DOI 10.1109/TII.2019.2909305

[2]Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 7, pp. 1693-1702, 2017.

[3]L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 8, pp. 1887-1899, 2018.

[4] X. Han, Y. F. Zhu. Prediction of Residual Fatigue Life of In-service Gears[J]. Journal of Wanxi University, 2003, 19(2): 85-87.

[5]R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213-237, 2019.

[6]Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic Sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," IEEE Transactions on Industrial Electronics, vol. 66, no. 5, pp. 3814-3824, 2019.

[7]Gebraeel N, Lawley M, Liu R, et al. Residual life predictions from vibration-based degradation signals: a neural network approach[J]. IEEE Transactions on Industrial Electronics, 2004, 51(3):694-700.

[8]Graves A, Mohamed A R, Hinton G. Speech Recognition with Deep Recurrent Neural Networks[J]. 2013, 38(2003):6645-6649.

[9]Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 2002, 5(2):157-166.

[10]Hochreiter S, Schmidhuber J. Long short-term memory.[J]. Neural Computation, 1997, 9(8):1735-1780.

[11]M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network," in Aircraft Utility Systems (AUS), IEEE International Conference on, 2016, pp. 135-140: IEEE.

[12] A. Elsheikh, S. Yacout, M.-S. Ouali, Bidirectional handshaking LSTM for remaining useful life prediction, Neurocomputing, 323 (2019) 148-156.

[12]K. B. Charbonneau and O. Shouno, "Neural Trajectory Analysis of Recurrent Neural Network In Handwriting Synthesis," arXiv preprint arXiv:1804.04890, 2018.

[14]H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," Neurocomputing, vol. 284, pp. 171-176, 2018.

[15]P. Anderson et al., "Bottom-up and top-down attention for image captioning and visual question answering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6077-6086.

[16]L. Wu, Y. Wang, X. Li, and J. Gao, "Deep attention-based spatially recursive networks for fine-grained visual recognition," IEEE transactions on cybernetics, no. 99, pp. 1-12, 2018.

[17]X. Ran, Z. Shan, Y. Fang, and C. Lin, "An LSTM-Based Method with Attention Mechanism for Travel Time Prediction," Sensors, vol. 19, no. 4, p. 861, 2019.

[18]T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection," Neural networks, vol. 108, pp. 466-478, 2018.

[19]A. Najafi, A. Joudaki, and E. Fatemizadeh, "Nonlinear dimensionality reduction via path-based isometric mapping," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 7, pp. 1452-1464, 2016.

[20]P. Nectoux et al., "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in IEEE International Conference on Prognostics and Health Management, PHM'12., 2012, pp. 1-8: IEEE Catalog Number: CPF12PHM-CDR.