Chapter 55
Rolling Bearing Degradation State Prediction with Deep Fusion Feature

Hao Chen, Niaoqing Hu, and Lun Zhang

Abstract In industrial manufacturing, the safety and reliability of rolling bearing play an important role for the normal operation of equipment. Therefore, it is of great significance to predict the future health state of rolling bearings accurately. With the increasing complexity of mechanical equipment and its external environment, data-driven prognostic methods gain popularity thanks to the advanced sensor technology and massive computer power. It is critical to find a stable and monotonous degradation feature for the data-driven prognostic method. However, the most of traditional degradation features do not fully mine the bearing state information contained in the sensor data, this paper proposes a data-driven prognostic method based on long-short-term memory network (LSTM) with deep fusion feature. This method extracts the shallow features from the raw data, then takes the shallow features as the input of Deep Sparse Auto-encoder (DSAE) to extract deep degradation features, and finally uses LSTM to predict the bearing future degradation state. In this paper, experiment was performed using the IEEE PHM 2012 Data Challenge Bearing Dataset. The results show that the deep features are more sensitive to early faults and have better predictability, and LSTM performs well in terms of short-term health state prediction for bearings.

55.1 Introduction

Rolling bearing is one of the key components of industrial equipment, it mainly plays a role in supporting and reducing friction. Many studies have shown that the failure of most rotating machinery are mainly caused by bearing failure [1].

H. Chen · N. Hu (✉) · L. Zhang
Laboratory of Science and Technology on Integrated Logistics Support, National University of Defense Technology, Chang Sha, China
e-mail: hnq@nudt.edu.cn

H. Chen · N. Hu · L. Zhang
College of Intelligence Science and Technology, National University of Defense Technology, Chang Sha, China

Therefore, accurately predicting the degradation state and remaining useful life (RUL) is an important part of Condition Based Maintenance (CBM) for bearing, which can save economic costs and avoid serious industrial accidents. Up to now, how to get a good degradation feature is one of the challenges for rolling bearing prognostics [2].

Degradation feature, also known as health indicator, is the important parameter reflecting the degradation state of bearings. In general, a good degradation feature should have the characteristics of uniform monotonicity, obvious trend and low volatility. Common degradation feature include time-domain statistics feature, frequency-domain statistics feature, entropy and so on. However, in the process of bearing degradation, the type of fault may change, moreover, several different fault types can coexist in the same degradation stage, and these domain statistics features often sensitive to the only one fault type, which cannot better reflect the whole life degradation process of the bearing [3]. In order to get the global feature which can describe the whole life degradation process of the bearing, feature fusion is a common method.

General feature fusion methods include principal component analysis (PCA), Self-organizing map (SOM) and so on. Dong and Luo used PCA to merge the raw features and reduce the dimension, and then get a more sensitive degradation features [4]. Huang et al. used SOM method to integrate 6 vibration features and successfully extracted consistent and practical degradation indicators [5]. However, the above two methods also have the following shortcomings: PCA is a linear mapping and cannot handle the nonlinear relationship between different variables. SOM network maintains the topological continuity of data, but the basic SOM network has the disadvantages of fixed and nonadjustable network structure, and local information will disappear in the case of data dimensionality reduction.

In recent years, with the development of large-scale storage facilities and sensor networks, a large number of industrial equipment condition monitoring data can be obtained. Under the background of massive data, Auto-encoder, as an algorithm of unsupervised learning, can dig out the hidden information in the data and extract the abstract feature that accurately reflect the bearing degradation state of bearing [6]. RNNs have achieved good results in dealing with time series problems, and the condition monitoring data is essentially a time series. RNNs take both the historical data and the current data as input, and learn the dependence between sequence data through multiple trainings to predict the future development trend. However, basic RNNs often have problems of gradient disappearance and gradient explosion in the process of long sequence data training, and it is difficult to deal with the problem of long-term dependence. As a variant of RNN, LSTM network effectively solves these problems [7]. In this paper, a data-driven prognostic method based on LSTM with deep fusion feature is proposed. Firstly, multiple shallow features are extracted from the raw vibration data, and then these shallow features are fused using DASE method, so as to get the deep features that can represent the whole life process of bearing. Finally, the deep features are taken as the input of the LSTM network to predict the bearing future degradation state.

The rest of this paper is organized as follows: Sect. 55.2 introduces the basic theory of Auto-encoder and LSTM network. Section 55.3 introduces framework of the proposed method; Sect. 55.4 presents the experimental validations of the proposed method; the conclusion and future work will be given in the Sect. 55.5.

55.2 Theory

55.2.1 Auto-Encoder

Auto-encoder is an unsupervised learning neural network model, which minimizes the error between the reconstructed data and the raw data as learning goal, the error also called as loss function. The abstract representation of the raw data can be obtained by encoding. Auto-encoder can be divided into two parts: encoder and decoder. The basic auto-encoder is a two-layer neural network, and the simplified network structure of auto-encoder is shown in Fig. 55.1.
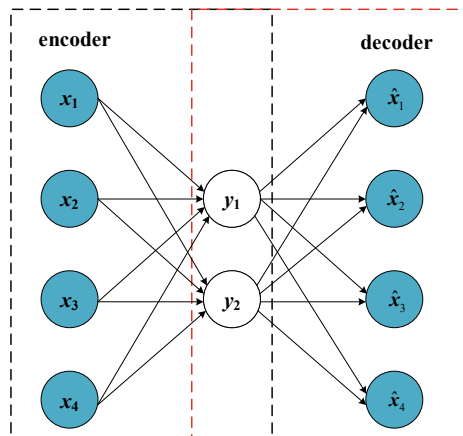
For the given input data $x = [x_1, x_2, x_3, \ldots, x_\mathbf{n}]$, the middle hidden layer output effective encoding feature, then the mapping relationship between the input layer and the $k\,th$ neuron in the hidden layer is as follows:

$$y_k = \sigma\left(W_1^k x + b^k\right) \tag{55.1}$$

Output reconstructed data are as follows:

$$\hat{x}_k = \sigma\left(W_2^k y + c^k\right) \tag{55.2}$$

Fig. 55.1 The structure of basic auto-encoder

After the output is obtained, the loss function can be calculated as follows:

$$L = \sum_{k=1}^{n} \left\| x^k - \hat{x}^k \right\|^2 + \lambda \|W\|_F^2 \tag{55.3}$$

where W represents the weight matrix, b, c all represent the bias term, λ is the regularization coefficient and σ denote the activation function.

In most cases, the auto-encoder with only two layers of neural network cannot achieve better data encoding feature. In order to obtain better data encoding, the deep network model is usually used. Stacked auto-encoder stack multiple auto-encoder together and extract feature layer by layer, it can get low dimensional, more representative deep features. Moreover, in order to obtain more abstract and representative compressed features, we can add a regularization item based on AE, also known as SAE, which can remain and extract as much as more information of the inputs. In this paper, L1–regularization is applied.

55.2.2 LSTM Network

The problem of gradient disappearance and gradient explosion often occurs in the training of long sequences with ordinary RNNs. LSTM network avoids the above problems by introducing a special structure named memory unit. As shown in the Fig. 55.2, an LSTM memory unit mainly includes: cell state and three gates (forget gate, input gate and output gate).

The LSTM memory unit have the ability to remove or add information to the cell state, it is regulated by the three gates. Forget gate determine what information is thrown away from the cell state, input gate determine what new information is input

Fig. 55.2 The basic structure of LSTM memory unit
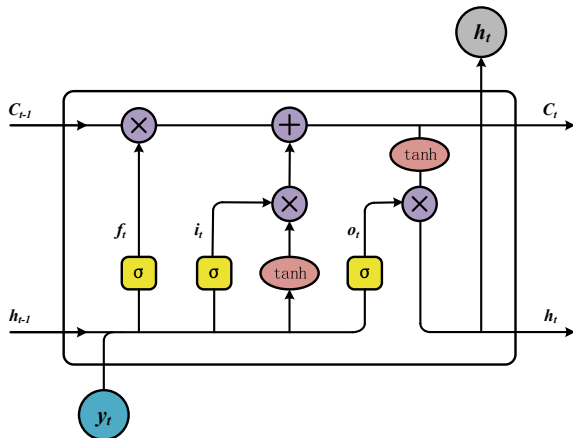
to the cell state, and output gate determine what will be output based on updated cell state.

At each time step t, the output of the three gates and the cell state are as follows:

$$
\begin{aligned}
f_t &= \sigma\left(\mathbf{W_f} * [h_{t-1}, y_t] + \mathbf{b_f}\right) \\
i_t &= \sigma(\mathbf{W_i} * [h_{t-1}, y_t] + \mathbf{b_i}) \\
C_t &= f_t * C_{t-1} + i_t * tanh(\mathbf{W_c} * [h_{t-1}, y_t] + \mathbf{b_c}) \\
o_t &= \sigma(\mathbf{W_o} * [h_{t-1}, y_t] + \mathbf{b_o}) \\
h_t &= o_t * tanh(C_t)
\end{aligned}
\tag{55.4}
$$

where f_t, i_t, o_t, C_t respectively represent the value of forget gate, input gate, output gate and cell state at t time step, h_{t-1} is the output values from the hidden layer at the previous time, $\mathbf{W_f}, \mathbf{W_i}, \mathbf{W_c}, \mathbf{W_o}$ denote weight matrix between the input layers and hidden layers, $\mathbf{b_f}, \mathbf{b_i}, \mathbf{b_c}, \mathbf{b_o}$ are the bias values of gate and cell state., and $*$ donates the product of two matrices, $[,]$ means that two vectors are concatenated to generate a new vector.

55.3 Methodology

The main work of this paper can be divided into three parts: shallow feature extraction, feature fusion and prognosis. Figure 55.3 summarize the flow of proposed method, some details described as follows:

- Vibration data preprocessing. General data preprocessing includes removal of outliers, noise reduction and so on. Considering that the vibration amplitude increases every moment in the degradation process, only noise reduction is
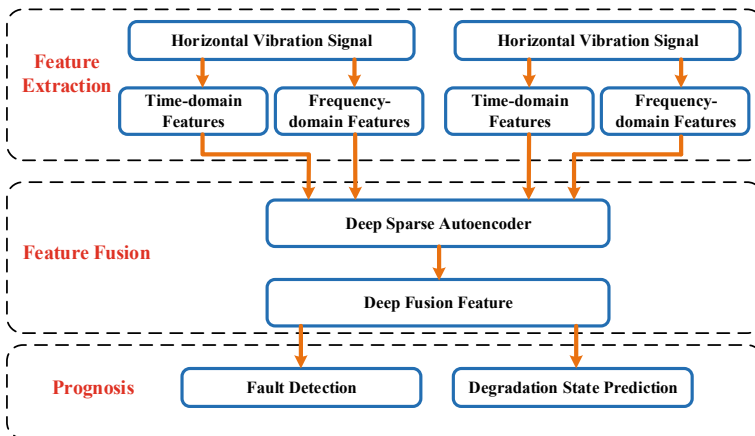

Fig. 55.3 The framework of the proposed method

carried out for the raw vibration data. In this paper, moving average smoothing method is used to eliminate the noise of the raw signal.

- Extract shallow features. Time-domain features and frequency-domain features are extracted from the preprocessed data.
- Feature fusion. Taking the extracted shallow features as the input of DSAE, the deep features with better monotonicity and obvious trend can be obtained.
- Prognosis. Detecting the early fault and taking the smoothed deep features as the input of the LSTM network to predict the bearing future degradation state.

55.4 Experiment and Analysis

55.4.1 Dataset Description

The PHM 2012 Data Challenge Bearing Prognostic Dataset used in this paper is provided by France FEMTO-ST Institute, which was obtained on PRONOSTIA platform for bearings accelerated life test [8].The experiment was divided into three working conditions. The horizontal and vertical vibration signal were collected by two acceleration sensors with the sampling frequency at 25.6 kHz, and 2560 samples (i.e. 1/10 s) are recorded each 10 s. When the amplitude of the vibration signal overpassed 20 g, it is defined as the bearing failure.

55.4.2 Results and Analysis

After the raw vibration signal is denoised, some features such as RMS are extracted from the pre-processed vibration signal. Figure 55.4 show RMS and kurtosis respectively.
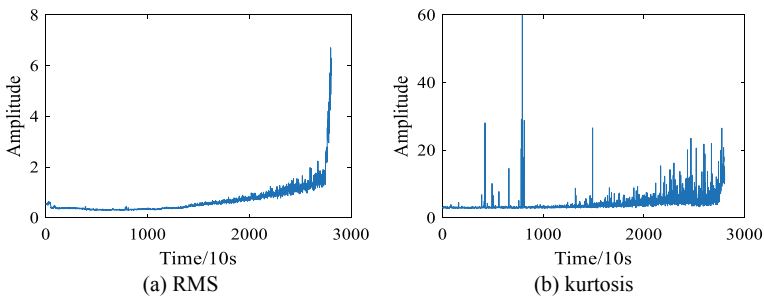

(a) RMS   (b) kurtosis

Fig. 55.4 RMS and kurtosis of bearing1_1

Table 55.1 Candidate shallow degradation features

Feature	Equation
Root mean square	$X_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{n} x_i^2}$
Variance	$X_{var} = \frac{1}{N}\sum_{i=1}^{n} (x_i - \bar{x})^2$
Square mean root	$X_{smr} = \left(\frac{1}{N}\sum_{i=1}^{n} \sqrt{|x_i|}\right)^2$
Mean frequency	$X_{fm} = \frac{1}{N}\sum_{i=1}^{n} y_i$ (y_i is the Frequency amplitude sequence)

As can be seen from Fig. 55.4, RMS can better reflect the degradation process of the bearing. This is because RMS represents the energy of vibration signal. When the bearing has a serious fault, the value of RMS will increase, but RMS is not sensitive to early fault. Kurtosis is sensitive to bearing fault, with the continuous degradation of the bearing, the impact component in the signal increases obviously, but the monotonicity is not good due to the influence of fault impact signal, so it cannot well track the degradation process of bearing. A good degradation feature should have the characteristics of good monotonicity, obvious trend, and strong robustness. Therefore, some bad degraded features should be removed before feature fusion. In this paper, we select Root mean square (RMS), variance, square mean root and mean frequency of horizontal and vertical signals total 8 features, as shown in Table 55.1.

Before the shallow features are input into DSAE to extract the deep features, the shallow features are normalized with Min-max normalization. In this paper, the DSAE model consists of three auto-encoders. The first layer is the input layer, and the number of neuron nodes is 8 as well as the input dimension. The number of neuron nodes in per layer is $[8, 40, 10, 1, 10, 40, 8]$. L1 regularization is adopted in the input layer, where the number of nodes in the second layer is more than that in the first layer to extract more sparse features. The activation function of the last layer is Tanh function, and the other layers' activation function is the ReLU function. The bath size is 100 and the epoch of network iterations is 200. Adam optimization algorithm is used in the training. The loss function is defined as the root mean square error (RMSE), and the formula of the RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{55.5}$$

where y_i the denotes real value, $\hat{y}_i$ denotes the predicted value, and n denotes the length of data sequence.

Figure 55.5 show the bearing1_1 deep features under working condition 1 and bearing2_1 deep feature under working condition 2. It can be seen from Fig. 55.5
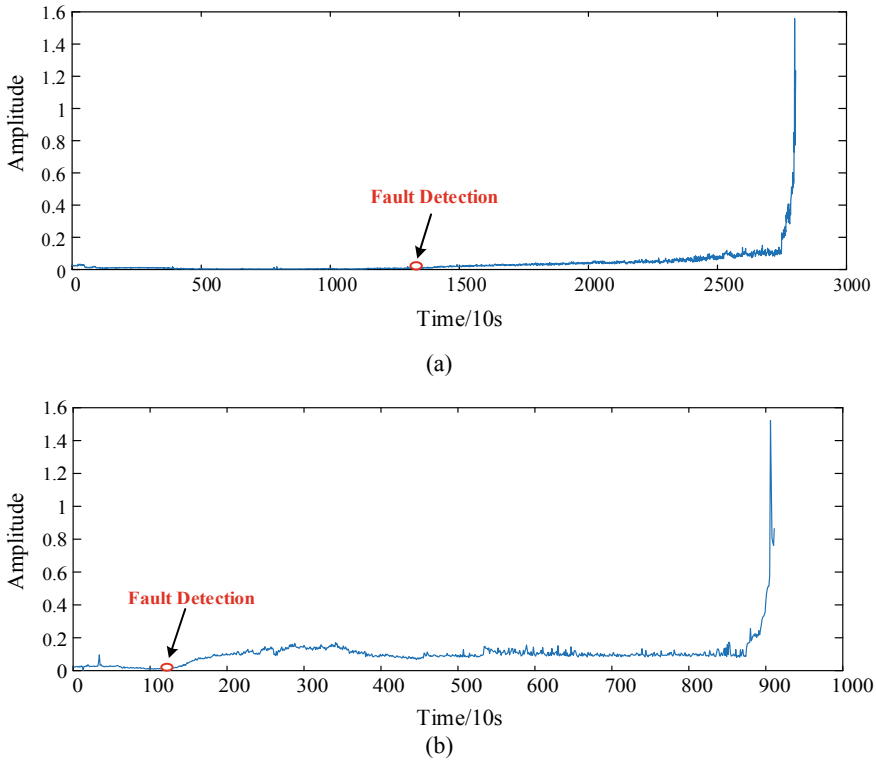
**Fig. 55.5** The deep fusion feature: **a** bearing 1_1; **b** bearing2_1

that the deep bearing features extracted by DSAE are less fluctuation than shallow features such as RMS and have obvious trend, besides, it is more sensitive to early fault. However, the deep feature extracted by DSAE is not very sensitive to the early fault as well as RMS under working condition 3, because the working condition 3 of heavy load and high speed, the process time from bearing early fault to failure is relatively short.

After obtaining the smoothed deep features, LSTM network is used to predict the future degradation of bearings. In this paper, Tensorflow framework is used to build the LSTM network model. The LSTM model has 2 hidden layers, the numbers of 1st hidden layer nodes are 160 and 2nd hidden layer nodes are 100. Input time step is set to 50. Adam optimization algorithm is used in the training, the number of network iterations is 100 and learning rate is 0.006. At the same time, in order to prevent overfitting during the training process, the dropout layer is added to the model, the dropout rate is set to 0.4. The loss function is also the RMSE.

Figure 55.6 show the prediction of bearing1_3 degradation state with prediction step of 10 and 20. In this paper, the LSTM model is trained with the data of bearing1_1 to predict the degradation state of bearing1_3. As can be seen from
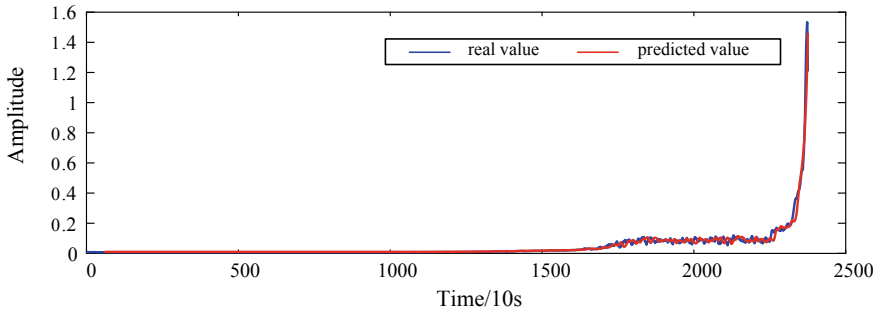
(a) The prediction step is 10
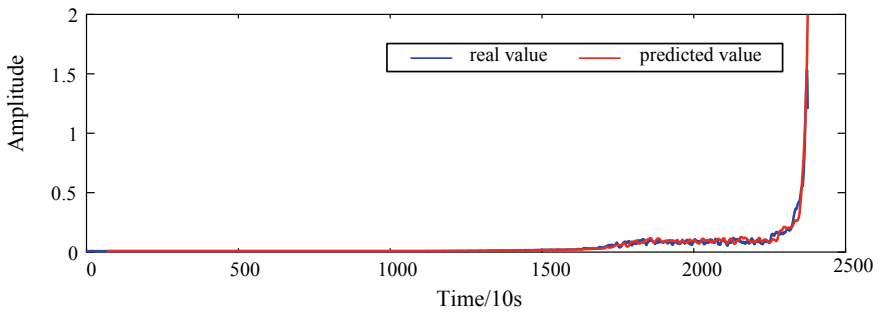
(b) The prediction step is 20

Fig. 55.6 The prediction of bearing1_3 degradation state

Fig. 55.6, the LSTM model can predict the degradation state of bearing well, it's worth noting that the prediction accuracy will decrease with the increase of the prediction step.

55.5 Conclusion

A data-driven bearing prognostic method based on LSTM with deep feature is proposed in this paper. Considering that a shallow degradation feature cannot make full use of the hidden state information in the data, this method integrates multiple shallow degradation features such as RMS through DSAE, so as to obtain more effective deep degradation features which can better characterize the whole life degradation process of bearings. Finally, LSTM network can be used to realize the short-term prediction of bearing degradation state. And it is still a tedious task to extract multiple features manually. Next, we will consider using the auto-coder to extract features directly from the raw data. And in this work, we have tried to use

LSTM for long-term prediction of bearing, but the result is not good, how to realize accurate long-term prediction of bearing is our focus work in the future.

References

1. Williams, T., Ribadeneira, X., Billington, S., Kurfess, T.: Rolling element bearing diagnostics ın run-to-failure lifetime testing. Mech. Syst. Signal Process. **15**, 979–993 (2001)
2. de Azevedo, H.D.M., Araújo, A.M., Bouchonneau, N.: A review of wind turbine bearing condition monitoring: State of the art and challenges. Renew. Sustain. Energy Rev. **56**, 368–379 (2016)
3. Zhao, L., Wang, X.: A deep feature optimization fusion method for extracting bearing degradation features. IEEE Access **6**, 19640–19653 (2018)
4. Dong, S., Luo, T.: Bearing degradation process prediction based on the PCA and optimized LS-SVM model. Measurement **46**, 3143–3152 (2013)
5. Huang, R., Xi, L., Li, X., Richard Liu, C., Qiu, H., Lee, J.: Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. Mech. Syst. Signal Process. **21**, 193–207 (2007)
6. Ren, L., Sun, Y., Cui, J., Zhang, L.: Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. J. Manuf. Syst. **48**, 71–77 (2018)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**, 1735–1780 (1997)
8. Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., Varnier, C.: PRONOSTIA: An experimental platform for bearings accelerated degradation tests (2012)