



Bearing performance degradation assessment using long short-term memory recurrent network

Bin Zhang^a, Shaohui Zhang^b, Weihua Li^{a,*}

^aSchool of Mechanical and Automotive Engineering, South China University of Technology, 381 Wushan Rd, Guangzhou, 510641, China

^bSchool of Mechanical Engineering, Dongguan University of Technology, Dongguan, 523808, China



ARTICLE INFO

Article history:

Received 26 August 2018

Received in revised form 13 December 2018

Accepted 20 December 2018

Available online 31 December 2018

Keywords:

Bearing
Degradation assessment
Health monitoring
Waveform entropy
Deep learning

ABSTRACT

Bearing is commonly used in rotating machinery, and it is significant to monitor bearing running states to ensure machine safety. Performance degradation assessment is an important work in bearing condition-based maintenance (CBM) and predictive maintenance. In this paper, a data-driven bearing performance degradation assessment method based on long short-term memory (LSTM) recurrent neural network (RNN) is proposed to comprehensively utilize the fault propagation information. Firstly, universal degradation simulation model based on vibration response mechanism is constructed for feature verification. A new proposed indicator "waveform entropy (WFE)" is developed and validated by this simulation model. Then waveform entropy and other conventional indicators are input into the LSTM RNN to identify the bearing running state, while particle swarm optimization method is applied to optimize the network structure parameters simultaneously. Experimental results demonstrate that LSTM RNN can effectively identify the bearing degradation states and accurately predict the remaining useful life.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Rolling element bearings withstand a variety of mechanical stress and thermal stress during the operation of machinery. According to [1], more than 40% of motor failures are related to bearing faults. Hence, accurate fault detection [2,3] and remaining useful life (RUL) prediction are rewarding to decrease the maintenance cost and reduce costly downtime. Many effective methods have been proposed and applied for bearing intelligent diagnosis [4–6], such as vibration analysis, temperature analysis and oil analysis. Among all these methods, vibration analysis [7–9] is the most convincing one and numerous indicators derived from vibration signals are implemented as criteria in traditional fault diagnosis.

Time-domain features (e.g., root mean square 'RMS' and kurtosis) are widely utilized to assess bearing performance degradation. Liao [10] extracted ten time-domain features as the input of the ANN and regrouping particle swarm optimization was utilized for the optimization of network structure. However, the fluctuation of conventional indicators affects the assessment of the bearing running state. Additionally, the fault threshold, especially for slight

fault detection, is hard to determined only by time-domain features. Frequency-domain indicators are also served as the criteria to identify bearing running state because of the development of the FFT algorithm. When the defect occurred, the corresponding fault frequency would be aroused, and the signature presented in a wide frequency band. Liao [11] picked up bearing fault frequency band energy and other time-domain indicators, and employed genetic programming to construct bearing degradation indicators for RUL prediction. Frequency-domain analysis can make the bearing defect more visualized, but it is not such efficacious when applied to the inner ring defect and the complex defect because of the complex spectrum in these cases. Another disadvantage is that frequency analysis requires expertise.

Data-driven methods (e.g., hidden markov model, support vector machine and artificial neural networks) have caught increasing attentions in past decades. Various indicators (time-domain, frequency-domain, time-frequency domain etc.) have been used to catch the practical fault information from high-dimension information. Chen [12] developed a more smooth indicator named relative root mean square to avoid the influences of outliers. Selak [13] proposed a condition monitoring and fault diagnostic system for hydropower plants, in which support vector machine was applied for fault identification. Li [14] used adaptive Stochastic Resonance (ASR) to enhance bearing fault feature in incipient stage and construct a quantitative evaluation system. Li [15] improved locality preserving projection by simultaneous

* Corresponding author.

E-mail addresses: jog_back@foxmail.com (B. Zhang), zhangsh@dgut.edu.cn (S. Zhang), whlee@scut.edu.cn (W. Li).

analyzing both the nearest and farthest samples of the feature space to enhance the fault classification and clustering performance. Chen [16] applied multiple two-layer sparse auto-encoder (SAE) neural networks for feature fusion, and the fused feature vectors were used to train deep belief network (DBN) for bearing fault classification. Qian [17] implemented recurrence quantification analysis (RQA) to extract features, which were served as the input of an auto-regression model to predict bearing states, and then Kalman filter was utilized to optimize the prediction. Chen [18] used a visualized data partition method based on spectral decomposition to avoid the misjudgment caused by poor quality datasets with outliers. Benkedjouh [19] processed vibration features by isometric feature mapping technique (ISOMAP), then these features were input into support vector regression (SVR) model to predict bearing RUL. Zhao [20] adopted Gated Recurrent Unit (GRU) Networks to learn representation of extracted features. However, the fusion feature generated by dimensionality reduction algorithms often has weak physical meaning. Moreover, the validity of these different indicators should be verified by a universal degradation model to avoid the influence of individual data sets.

Beyond feature extraction and fusion, data-driven model could summarize the characteristic of different degradation stage and catch the same characteristic hidden in testing samples, which paves another path to RUL prediction. Huang [21] extracted minimum quantization error (MQE) based on self-organizing map as an indicator to estimate bearing RUL. Compared with the L10 life, this method give a new bearing RUL guidance. But the RUL prediction error of 14 test bearings are still around 20%. Yu [22] proposed negative log likelihood probability (NLLP) based on gaussian mixture model. Compared with the traditional indicators, NLLP has a good monotonous trend, which is helpful to judge the bearing degradation degree. But this indicator is insensitive to bearing incipient fault. Ben [23] proposed a new indicator named root mean square entropy estimator, which was used as the input of simplified fuzzy adaptive resonance theory map neural network (SFAM) to identify the bearing fault degree. The indicators of two consecutive time points were utilized. This study transformed degradation assessment to classification tasks and the classification rate reached 74.2%. Its disadvantage is that the SFAM model

could not discover the relationship of sequential inputs. Guo [24] adopted recurrent neural network to construct a fusion indicator named RNN-HI. A double exponential model was applied for RNN-HI curve fitting and prediction. The mean of testing RUL error is 32.48%. However, bearing usually running smoothly before fault occurrence, this method linearly labeled the whole degradation from 0 to 1 to obtain RNN-HI, which was inconsistent with the actual degradation process.

Although many researchers have achieved great progress in the field of bearing performance degradation assessment, there are still some problems need to be resolved:

- (1) Automatic detection of the fault occurrence. The defect occurrence is the start of degradation, which is important to predict the RUL. According to the detection of fault occurrence, the fault propagation will be tracked and assessed by the model. However, for most existing methods, the fault occurrence is determined by experts' judgments.
- (2) A health indicator should have clear physical meaning and exhibits a monotonically trend during degradation. And the effectiveness will be undermined if the indicators only perform well in specific conditions. So, its robustness and effectiveness need to be verified.
- (3) The usage of historical data. For bearings, the degradation is a continuous process, the fault occurring, propagating, developing and damaging information should be utilized as a whole life data for construction of the assessment model.

In this paper, a data-driven prognosis method based on LSTM recurrent network is proposed. Firstly, we construct a bearing degradation indicator, called waveform entropy, to represent bearing running states. The effectiveness of the proposed indicator is further verified by a universal degradation simulation model based on vibration response mechanism. After feature verification and selection, bearing states are identified by LSTM recurrent network, whose structure parameters are simultaneously determined by particle swarm optimization (PSO) method. In order to catch reasonable samples and better training accuracy, a pre-training of fault occurrence point identification is implemented. Experimental results demonstrate that LSTM

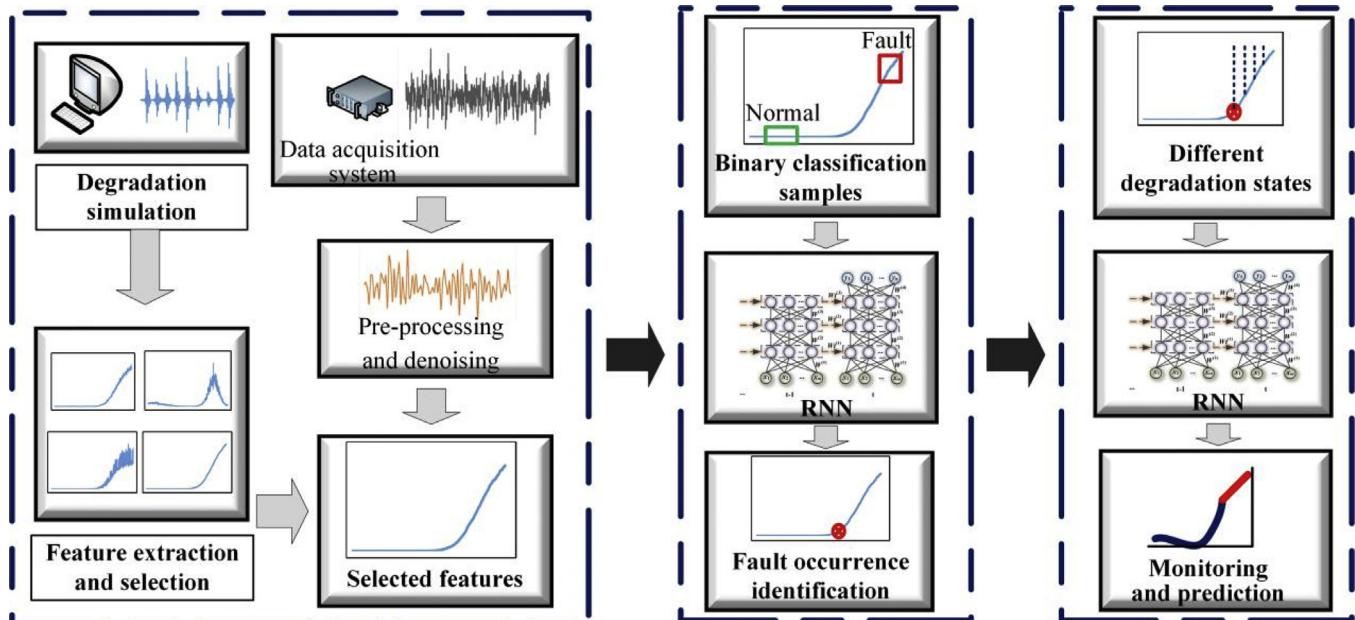


Fig. 1. Flowchart of proposed bearing performance degradation assessment method.

networks can effectively identify the bearing degradation degree and accurately predict the RUL. The main contributions of this study can be concluded as follows:

- 1) A dimensionless indicator called waveform entropy is proposed to represent bearing running states, which exhibits better robustness and stronger monotonicity.
- 2) Considering the bearing vibration mechanism, a degradation simulation model is constructed for feature verification and selection. Different simulation conditions can be set to test the performance of each feature.
- 3) An automatic detection method is proposed to identify the fault occurrence, and thus the degradation process can be segmented into different stages according to time stamps.
- 4) LSTM recurrent network is employed for bearing degradation assessment and its structure parameters are optimized by PSO method. The effectiveness of the proposed method is validated by run-to-failure experimental datasets.

The remainder of this paper is organized as follows: Section 2 introduces the process of the proposed bearing performance degradation assessment. Section 3 describes the feature construction and verifies its effectiveness. Section 4 introduces the structure of LSTM recurrent neural network and its parameters determination process. Section 5 presents the experimental results and analyses. Finally, Section 6 concludes the whole works.

2. The RUL prediction procedure of the proposed approach

In this work, a bearing degradation assessing method based on LSTM recurrent neural network is proposed, which can be divided into three steps as follows:

- (1) Degradation indicator selection. In this step, a degradation indicator named waveform entropy is developed. To verify the performance of different indicators, a bearing degradation simulation model based on vibration response mechanism is proposed. The robust features would be selected as input of assessment model.
- (2) Automatic detection of bearing fault occurrence. The normal state of bearing has a long duration and should be distinguished from fault state when making training samples. Prognosis of bearing fault occurrence is a pre-training of final classification, which can improve the rationality of labeling samples.
- (3) Classification of degradation stages. The degradation process is further divided into several stages after fault occurrence identification. Then LSTM recurrent network model was built to recognize the degradation level through classification. Each class and its possibility are used to calculate a degradation index, then an exponential degradation model is applied to predict the degradation trend.

3. Feature extraction

Feature extraction is critical to degradation assessment because bearing vibration signal contains rich hidden information. Many time-domain features are adopted to explore the relationship between the bearing running states and vibration signals. Among them, RMS and kurtosis are most common features, since they can effectively reflect the real-time change of the bearing vibration. However, the fluctuation caused by the uncertainty of vibration has a great influence on the estimation of the bearing degradation. Therefore, further processing of the traditional features, e.g., feature fusion or statistical analysis, is needed to find a good indicator.

3.1. Waveform entropy feature

For a single discrete random vector X , its entropy can be calculated by

$$E = \frac{1}{M} \sum_{i=1}^M p(x_i) * \ln(p(x_i)) \quad (1)$$

Where x_i represents an element in X and its probability mass function is $p(x_i)$. M is the total number of elements in variable X .

When the bearing defect occurs, the vibration shows an increasing trend with the fault developing. Despite discontinuous saltus may occurs, the average vibration energy during this period has a great correlation with bearing degradation. Entropy is the measurement of system average uncertainty and chaos. From the prospect of statistics, the vibration entropy in a period can better depict the real bearing state. Based on this hypothesis, it is reasonable to construct an entropy feature to reflect the degradation degree.

For time-domain features, the dimensional indicators are related to not only the state of the equipment but also the operating conditions (speed and load), whereas the dimensionless indicators are related to the state of the machine and not sensitive to the change of load and speed. Therefore, we prefer to construct a dimensionless indicator, such as some kind of entropy features.

Waveform factor is defined as the ratio of RMS to the average of absolute amplitude, which reflecting the deviation from the average. And it is not sensitive to operation conditions. Inspired by Ben [23], we construct a dimensionless indicator called Waveform Entropy(WFE), defined by:

$$WFE_t = \frac{1}{M} \sum_{i=0}^{M-1} W_{t-i} * \log(W_{t-i}) \quad (2)$$

Where WFE is the waveform entropy, W_t is the waveform factor of vibration signals at time t , and M is the length of sliding window.

As it can be seen, the calculation of WFE do not need signals decomposition and complex parameter setting compared with

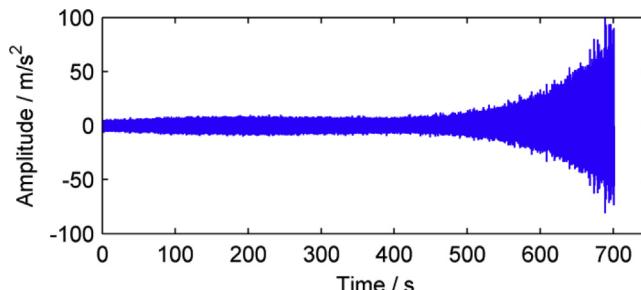


Fig. 2. Simulated bearing degradation signals of small-noise condition.

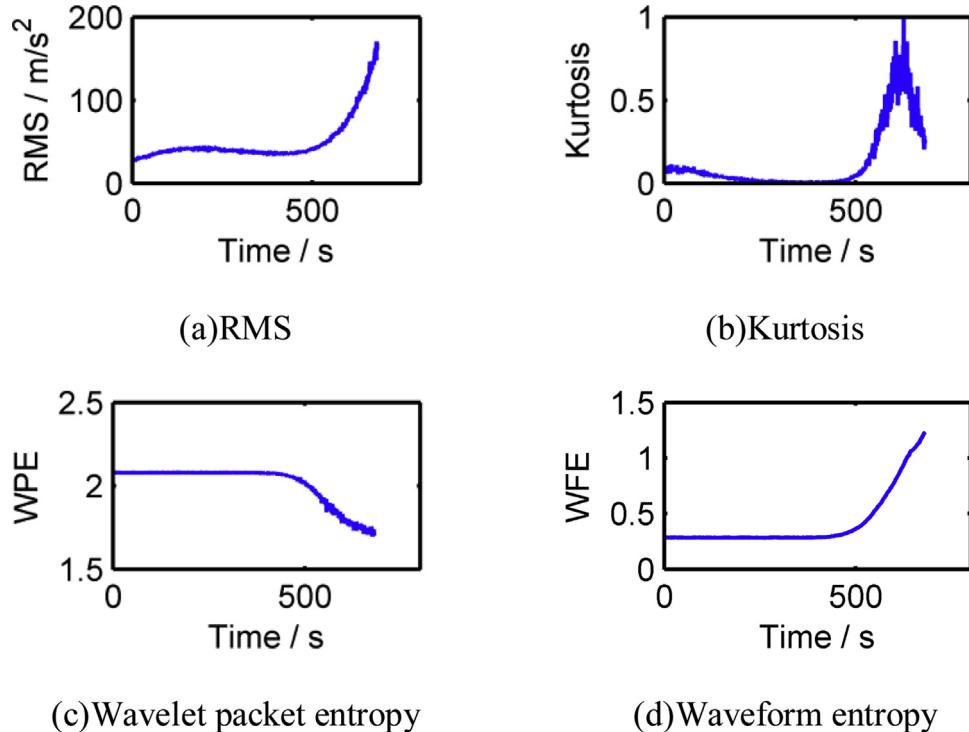


Fig. 3. Bearing degradation indicators of small-noise condition simulation.

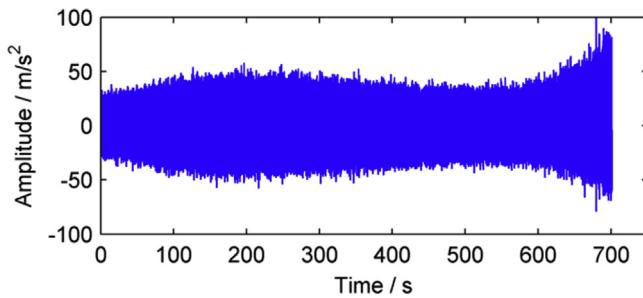


Fig. 4. Simulated bearing degradation signals of big noise condition.

time-frequency domain features. ‘Waveform entropy’(WFE) is the local mean of logarithmic vibration energy, which has clear physical meaning of the measure of vibration strength. Since waveform factor of vibration signals may be greater than 1, applying log function can keep ‘WFE’ from variating drastically.

3.2. Degradation simulation analysis

To compare the performance of different indicators, a bearing degradation simulation model are constructed according to vibration response mechanism [25]. This vibration signals mainly contains three components, which are vibration response signals, environment noise and system noise. The frequency component of vibration response signals depends on the fault type and operating conditions, and the amplitude is related to degradation degree. In this simulation, the components except for vibration response are regarded as noise components. Environmental noise is fixed-size noise in mechanical equipment installation position, which won't produce very big change. System noise is a vibration interference, which may come from inside the equipment and would increase as bearing damage becoming severe.

Combined with the exponential degradation model and the bearing fault vibration response mechanism, bearing degradation

signals could be simulated by

$$f(t) = \lambda_1 * \exp(\lambda_2) * x(t) + \eta_1(t) + \lambda_3 \eta_2(t) \quad (3)$$

$$x(t) = \sum_{i=1}^I \sum_{j=1}^J A_{ij} \exp \left[\frac{-2\pi\zeta_j}{\sqrt{1-\zeta_j^2}} f_{dj}(t - \tau_i - iT) \right] \sin \left[2\pi f_{dj}(t - \tau_i - iT) \right], \\ t \geq \tau_i \quad (4)$$

In Eq. (3), $f(t)$ is the final simulation signals, λ_1, λ_2 are two parameters which reflect vibration amplitude variation. λ_3 is set to describe the influence of system noise. $\eta_1(t)$ and $\eta_2(t)$ are Gaussian white noise. $x(t)$ is bearing damage vibration response component.

In Eq. (4), $f_{dj}, \zeta_j, j = 1, 2, \dots, J$ represent the corresponding system frequency and damping ratio under different modes, only the first few components are considered; A_{ij} represents the amplitude of j_{th} system frequency upon the i_{th} impulse, which is set to be a random value in range [0,0.5]. T represents the theoretical cycle of the impulse. τ_i represents the difference between theoretical cycle and real impulse time (Fig. 1).

Four working condition are set for simulation, which include constant speed with small-noise, constant speed with large-noise,

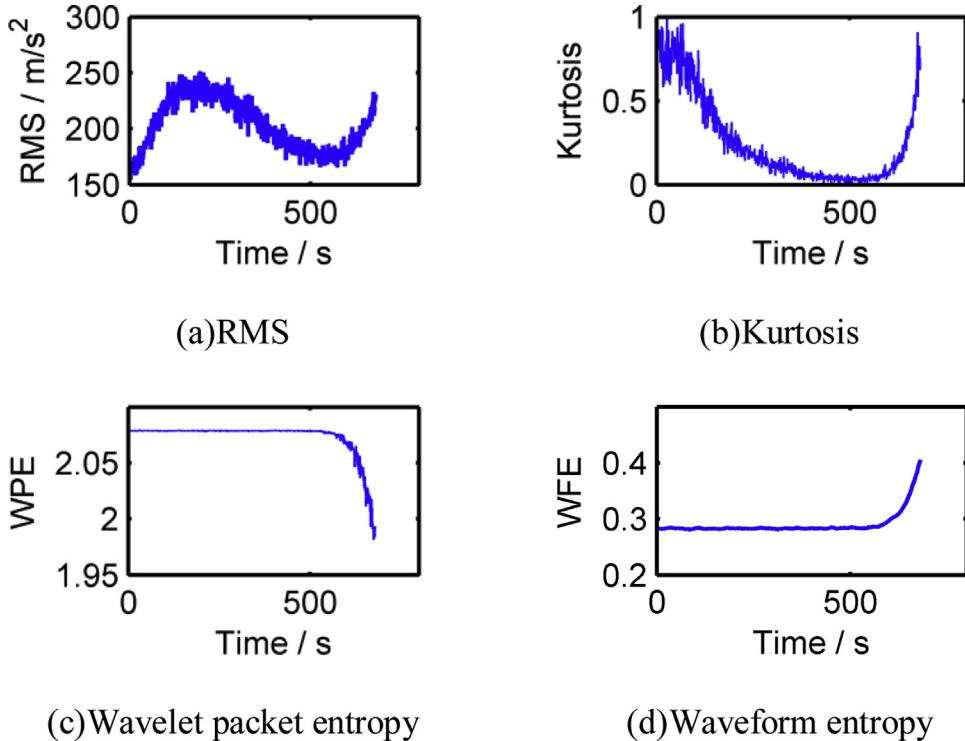


Fig. 5. Bearing degradation indicators of big noise condition simulation.

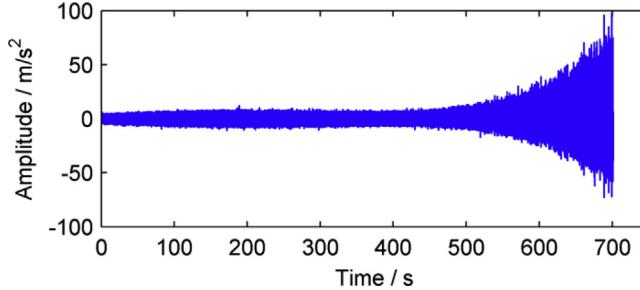


Fig. 6. Simulated bearing degradation signals of rotation fluctuation condition.

speed fluctuation without noise and speed fluctuation with large-noise. In each condition, we compare the performance of different indicators to validate the effectiveness. For comparison, bearing parameters are set the same as Rexnord ZA-2115 and rotation speed is 2000 rpm. Thus BPFO, BPFI and BSF are respectively 236.4 Hz, 396.9 Hz and 140.0 Hz. Assuming that the first two system frequency components are aroused when bearing damage occurred, and these two frequency components are respectively 2000 Hz and 4000 Hz while the damping ratio are respectively 0.1 and 0.05. The sampling rate is set to 12 kHz and sampling length is 1 s. To simplify the model, $\eta_1(t)$ and $\eta_2(t)$ are set to the same Gaussian white noise, λ_3 is set to λ_2^2 .

The first working condition is small-noise condition. The whole life of simulation bearing is set to 700 s and signal-to-noise ratio is -15db in time 350 s. The Gaussian noise in time 350 s is applied as the environmental noise of whole bearing life, while system noise is λ_3 fold than environmental noise. In this working condition, the 700 segments of degradation data are showed as Fig. 2.

Different indicators are extracted from the simulation signals, including RMS, kurtosis, wavelet packet entropy (WPE) and waveform entropy. Db3 wavelet is applied to wavelet decomposition and the decomposition layer number is set to 3. The length of

sliding window of waveform entropy is set to 20. The result is showed in Fig. 3. In this small-noise working condition, all of the four indicators have a good performance in degradation assessment. RMS and WPE present a good increasing trend with only a litter fluctuation. Kurtosis is fitted to diagnosis the early fault for its peak value in this period. Compared with other indicators, the proposed WFE has better monotonicity and less fluctuation.

In the second working condition, environment noise and system noise are further increased. The signal-to-noise ratio is -30db in time 350 s, and the Gaussian noise is also applied as the environmental noise of whole bearing life. Other parameters are set as same as those under the first condition. In this big-noise working condition, the 700 segments of degradation data the four indicators are shown in Figs. 4 and 5.

As shown in Fig. 5, RMS and kurtosis are heavily interfered by the noise, which present a big value even in the normal stage, whereas the WPE and the WFE show good monotonicity with little fluctuation.

Under the third working condition, the fluctuation of rotation speed was considered. The rotation speed is set as a random value in [1900, 2100] rpm in each segment of 700 degradation data, which means that the rotation fluctuation range is 10%. And the

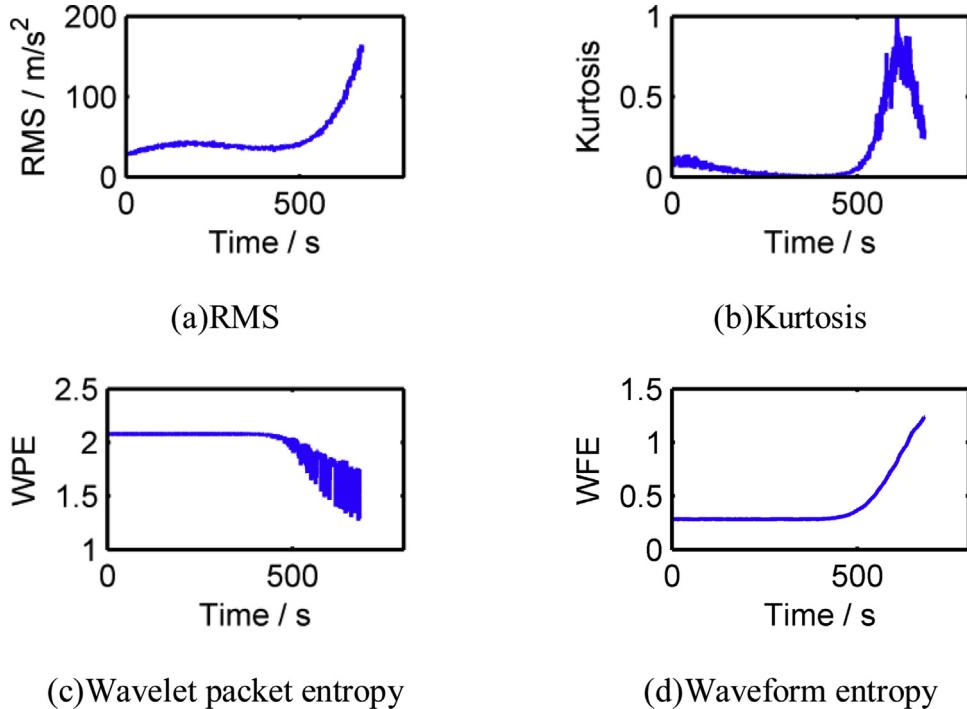


Fig. 7. Bearing degradation indicators of rotation fluctuation condition simulation.

700 segments of degradation data are showed as in Fig. 6 and the four indicators in Fig. 7.

In this condition, RMS, kurtosis and WFE perform as same as those under the first condition. But the WPE presents great fluctuation which may influence the diagnosis result.

Under the fourth working condition, both big noise and rotation fluctuation are considered, and the degradation process is showed as Fig. 8.

RMS, kurtosis, WPE and WFE are also calculated as showed in Fig. 9. It can be seen that the first three indicators suffer from severe interference under this working condition. However, WFE performs strong robustness and good monotonicity even with this couple interference.

Simulation results illustrate that WFE (waveform entropy) has advantages of strong robustness and good monotonicity, which is appropriate for degradation assessment and health monitoring.

4. LSTM recurrent network based assessing model

When assessing bearing degradation, utilizing the history data no doubt plays a positive role. A neural network, named as long short-term memory (LSTM) recurrent network, has memory

function and powerful series processing ability, which attracts much attentions recently.

The pioneer recurrent network are Elman networks and Jordan networks [26,27], in which the activation of nodes calculates not only the current input but also the previous return value. The recurrent structure endows the network with a superior sequential processing performance, but gradient disappearance problem comes together. With the increasing of time window length, historical information is progressively forgotten by the network. Hochreiter [28] proposed long short-term memory (LSTM) cells, adding three switches to hidden layer nodes. These switches, which named “gates”, determine the extraction ratio of the input and recurrent information, thus solving the problem of gradient disappearance.

LSTM-RNN achieves a really good performance in sequential data processing because of the recurrent feedback. For instance, Liwicki [29] applied LSTM recurrent network to handwritten digital recognition and achieved a state-of-art result. Sutskever [30] constructed two multi-layer LSTM networks for machine translation. In addition, LSTM have also been extensively applied in the language model, text generation, and speech recognition. Considering LSTM's ability of processing sequential data, we construct a degradation assessment model based on LSTM recurrent network.

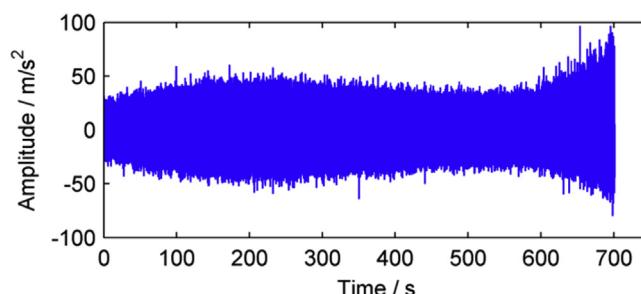


Fig. 8. Simulated bearing degradation signals of rotation fluctuation with big noise condition.

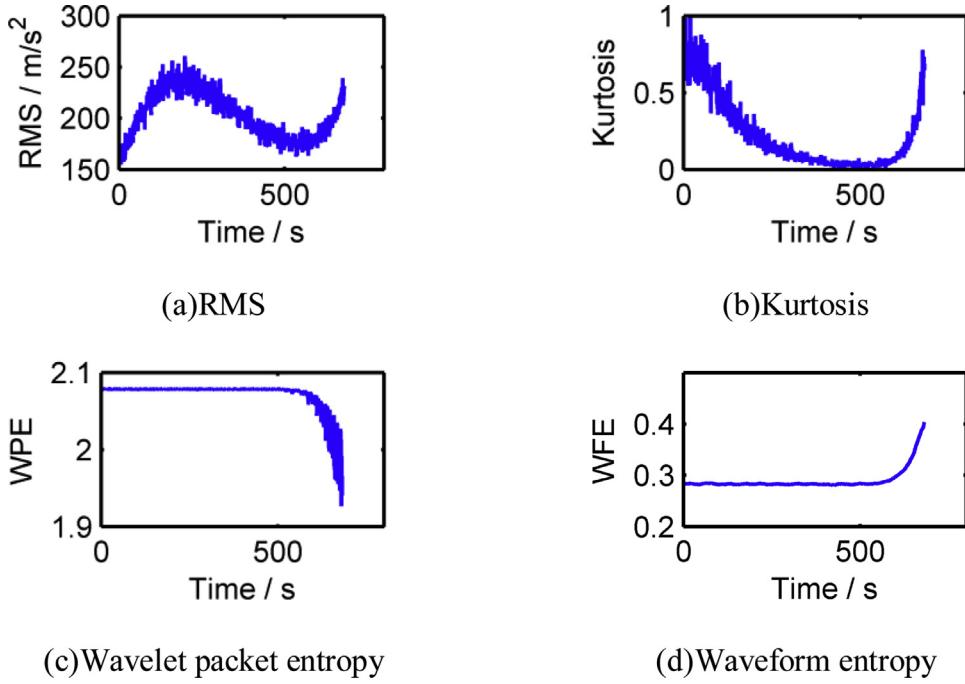


Fig. 9. Bearing degradation indicators of rotation fluctuation with big noise condition.

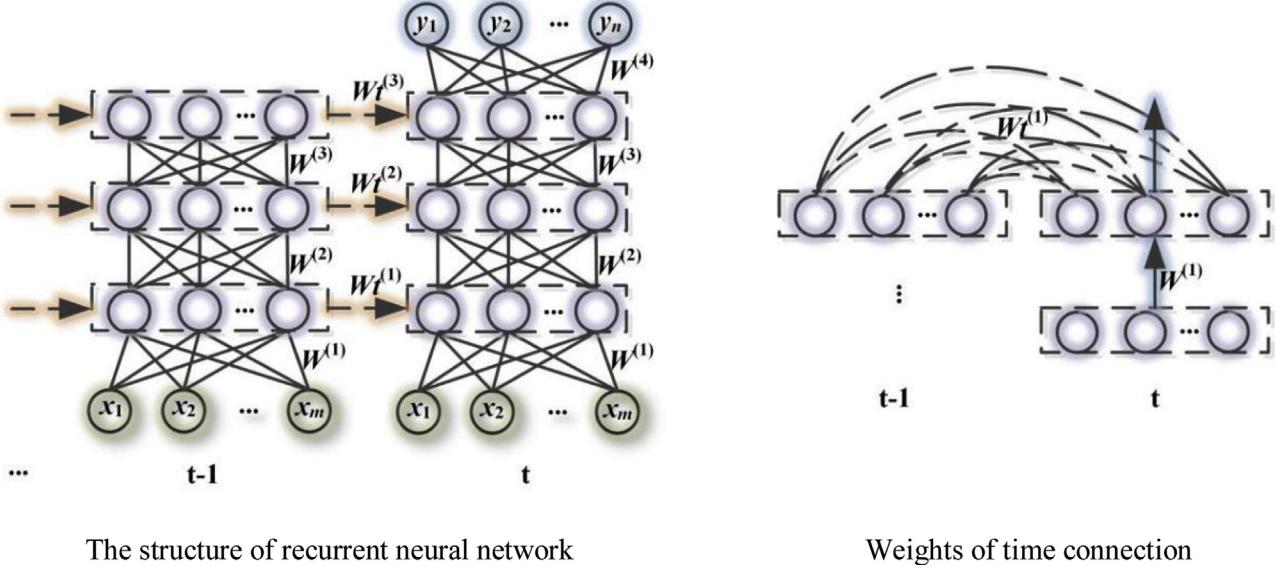


Fig. 10. LSTM recurrent network.

4.1. The structure of LSTM recurrent network

The recurrent network applied in this study is shown as Fig. 10, where x_1, x_2, \dots, x_m are the inputs of the network, y_1, y_2, \dots, y_n are the outputs. For example, the inputs can be the features extracted from vibration signals and the outputs can be the states of bearing. $W^{(1)}, W^{(2)}$ and $W^{(3)}$ are the weights of the layer connection, which are represented by solid lines. $W_t^{(1)}, W_t^{(1)}$ and $W_t^{(1)}$ are the weights of the time connection, which are represented by dashed lines.

Each hidden node in the above neural networks is a LSTM cell, whose structure is shown in Fig. 11. Where h_t represents the output of LSTM cell at current time t , s_t denotes the state of LSTM cell at time t , (g_s, i_s, f_s, o_s) represents the activation of the input and h_{t-1}

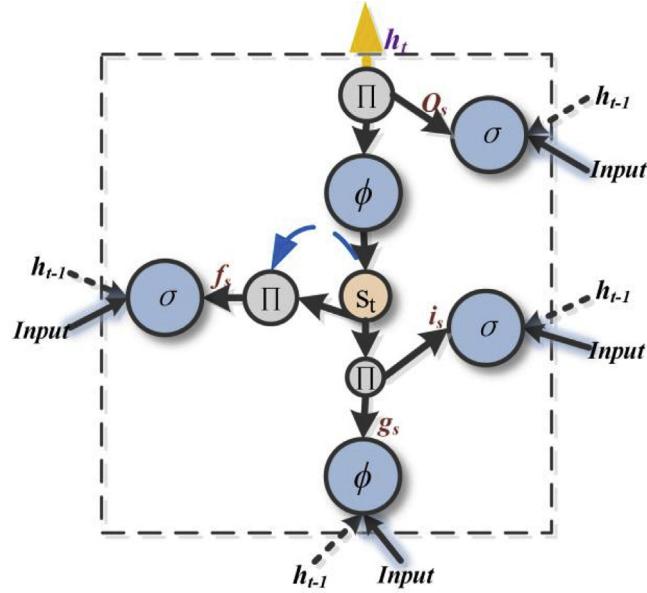
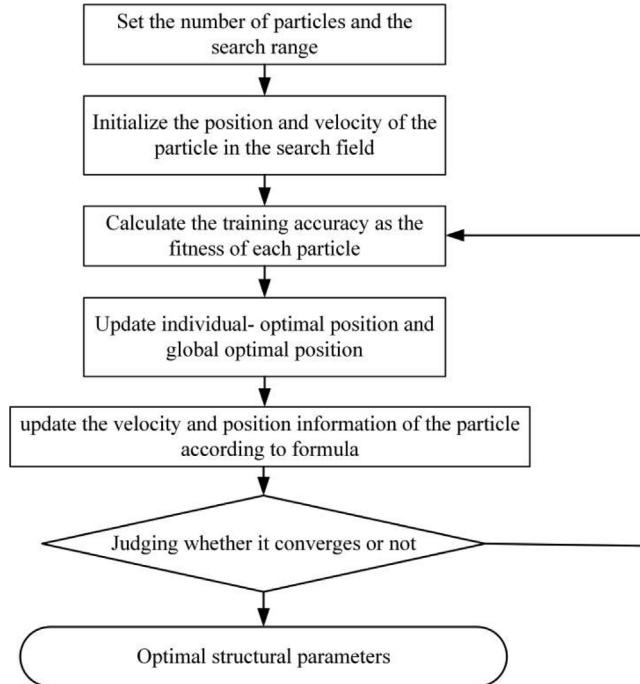
(i.e., activation of input nodes, input gate, forget gate and output gate), σ and ϕ are the activation function (sigmoid and tanh function), and ' \prod ' is the symbol for pointwise multiplication.

Once the input and the recurrent value h_{t-1} are weighting summed, the outputs of the LSTM recurrent network are calculated according to the following formula:

$$g_s = j(W_{gx} * x + W_{gh} * h_{t-1} + b_g) \quad (5)$$

$$i_s = \sigma(W_{ix} * x + W_{ih} * h_{t-1} + b_i) \quad (6)$$

$$f_s = \sigma(W_{fx} * x + W_{fh} * h_{t-1} + b_f) \quad (7)$$

**Fig. 11.** The structure of LSTM cell.**Fig. 12.** Structure hyperparameter optimization process using PSO.

$$o_s = \sigma(W_{ox} * x + W_{oh} * h_{t-1} + b_o) \quad (8)$$

$$s_t = g_s \otimes i_s + s_{t-1} \otimes f_s \quad (9)$$

$$h_t = \varphi(s_t) \otimes o_s \quad (10)$$

Where W_{gx} represents the “input-input node” weights matrix, W_{ix} the “input-input gate” weights matrix, W_{fx} the “input-forget gate” weights matrix and W_{ox} the “input-output gate” weights matrix,

respectively; W_{gh} , W_{ih} , W_{fh} and W_{oh} represent the weights matrix of recurrent value (h_{t-1}) and these four nodes (input node, input gate, forget gate, output gate) respectively; b_g , b_i , b_f , b_o are the bias of the summed weighted input, and \otimes represents pointwise multiplication.

4.2. PSO-based network structure parameter optimization

Like other deep neural networks, RNN is also sensitive to the setting of network hyperparameters, such as number of hidden layers, number of nodes per layer, time step and learning rate.

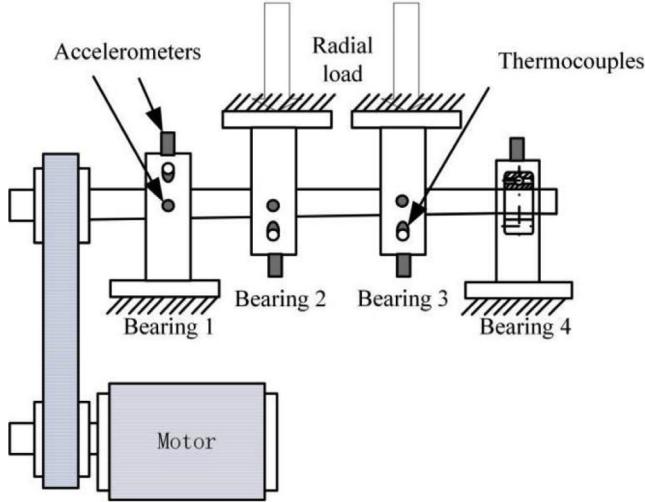


Fig. 13. Test rig [32].

Table 1

The running condition and fault type of run-to-failure bearings.

Bearing	Failure time	Failure type	Rotation speed /rpm	Load /lbs
TB1	2156	Inner race defect	2000	6000
TB2	2156	Roller defect	2000	6000
TB3	984	Outer race defect	2000	6000

Commonly, the deeper layers and the more hidden nodes will lead to better fitting. However, it will increasing computation cost and is also time consuming. Sometime, the coupling effects of different hyperparameters will influence the performance of the network.

To solve this problem, particle swarm optimization (PSO) is adopted in this study. PSO is a kind of multiple-parameter optimization algorithm, which is designed by simulating the foraging behavior of bird flock [31]. Assuming there is only one optimal solution in a certain district D , the position and velocity of m particles are initialized in this district. The position of particle represents a candidate solution, and the velocity determines the movement of the particle. The fitness of each particle could be calculated after initialization, and the personal best position P_{best} and the global best position G_{best} could be recorded. Then the position and velocity of m particles could be updated according to

the Eqs. (11) and (12):

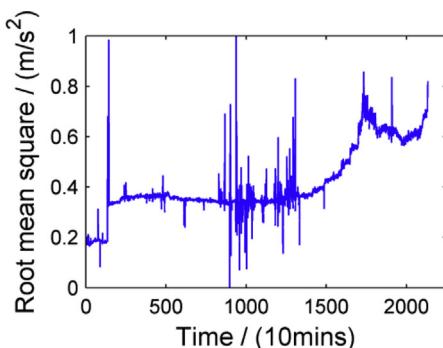
$$V_{id}^{k+1} = wV_{id}^k + c_1 r_1 (P_{best}^k - X_{id}^k) + c_2 r_2 (G_{best}^k - X_{id}^k) \quad (11)$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (12)$$

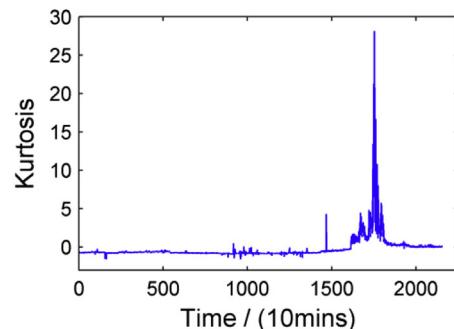
where V_{id}^{k+1} is the velocity of the i_{th} particle in the $(k+1)_{th}$ iteration, X_{id}^k is the position of the i_{th} particle in the k_{th} iteration, w is inertia factor, c_1 and c_2 are acceleration constants, r_1 and r_2 are the random value within $[0,1]$, d represents the d_{th} dimension of the position and velocity.

The fitness would be calculated again after updating the position, and the personal best position P_{best} and the global best position G_{best} are updated. If the maximum iteration is reached or the position error is smaller than preset value, the current position is the optimal solution. If fail to meet these two requirements, recalculate the fitness and best position.

The optimization process of structure hyperparameters is shown as Fig. 12. First, the number of particles m and the search range D are set, and the position and velocity of the particle are initialized in this range. All elements of V_{id}^k and X_{id}^k would be rounded to the nearest integer. Each set of parameters correspond to a particle, and the mean square error (MSE) of each training



(a) RMS



(b) Kurtosis

Fig. 14. Traditional features of TB2 bearing.

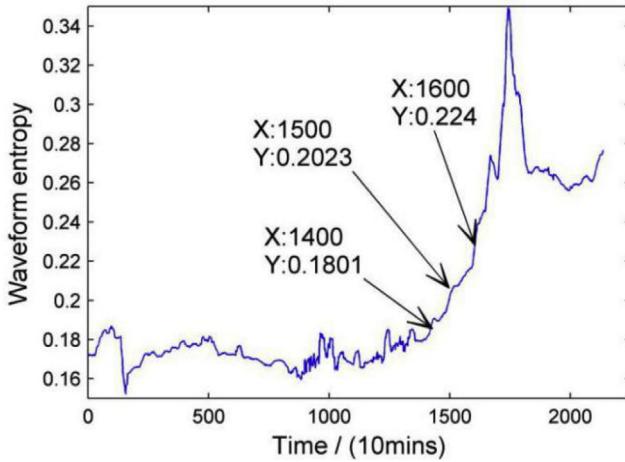


Fig. 15. Waveform entropy of TB2 bearing (ball fault).

Table 2
Bearing states binary classification samples.

Bearing	Failure time	Normal sample	Severe fault sample	Case 1 sample	Case 2 sample
TB1	2156	500-1000	1900-2100		✓
TB2	2156	500-1000	1700-2100	✓	✓
TB3	984	300-450	700-900	✓	

process of network can be set as fitness function of a particle. Then personal best position and global best position are updated according to the fitness of all particles. And the velocity and position of each particle can be updated by the new personal best

position and global best position. At last, optimal structure hyperparameters can be obtained (Fig. 13).

5. Bearing fault experiment

5.1. Bearing degradation state classification

The experimental data is obtained from NSF I/UCRC for Intelligent Maintenance Systems (IMS) [23]. The run-to-failure data sets are described in Table 1. Each file of the data sets consists of 20,480 points with a sampling rate of 20 kHz

Among these datasets, we select TB2 bearing for data analysis. RMS and kurtosis of TB2 bearing are shown as Fig. 14. The fluctuations and saltations of RMS make fault diagnosis more

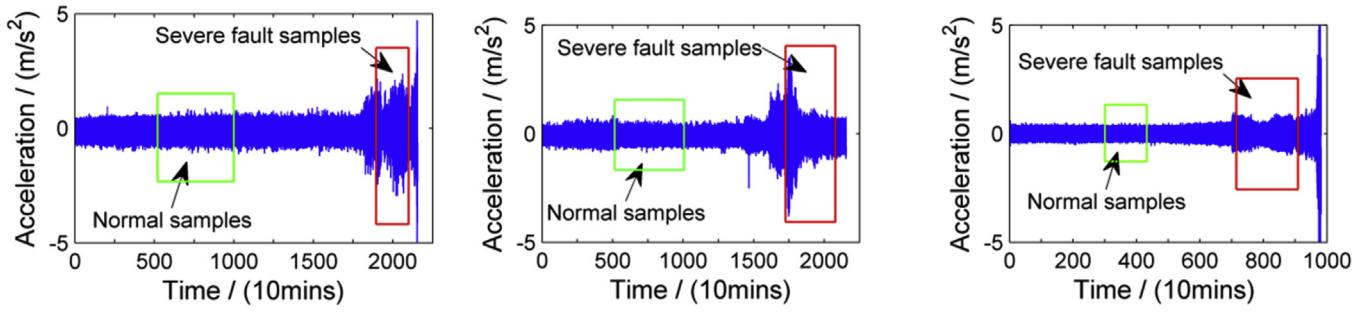


Fig. 16. Binary classification samples.

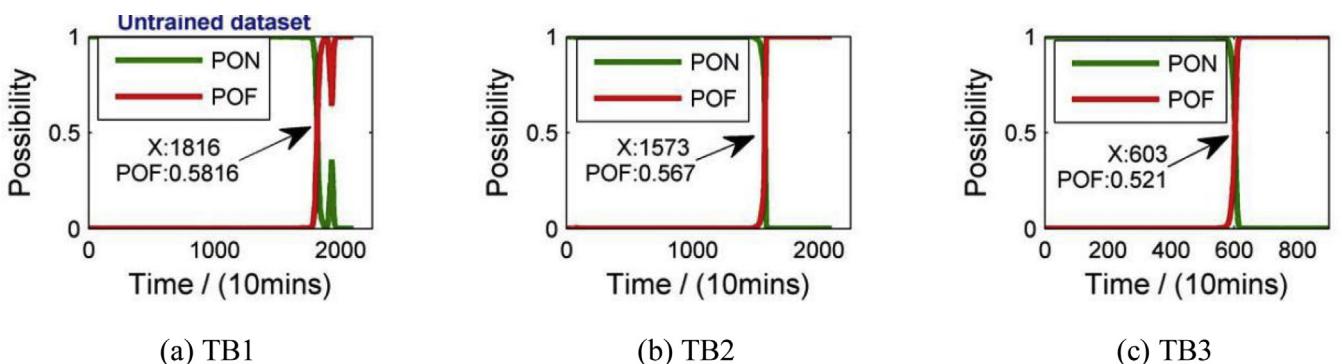


Fig. 17. Fault occurrence testing result in case 1.

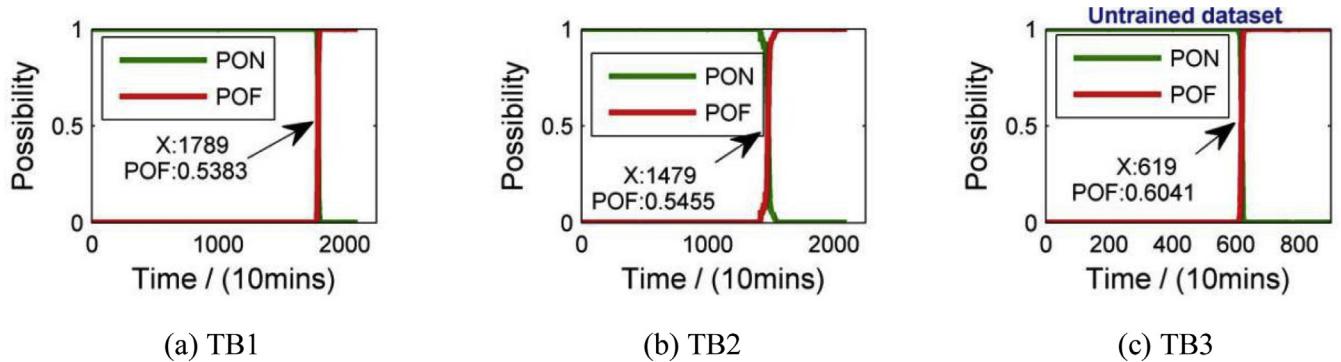


Fig. 18. Fault occurrence testing result in case 2.

Table 3
Classification results of methods.

Indicators	method	Training accuracy (%)	Test accuracy (%)
Waveform factor Kurtosis	LSTM recurrent network	89.428	78.455
Waveform factor Kurtosis Waveform entropy	LSTM recurrent network	94.968	93.117
Waveform factor Kurtosis Waveform entropy	BP networks	81.434	78.407
Waveform factor Kurtosis Waveform entropy	SAE	90.630	86.774
Waveform factor Kurtosis Waveform entropy	CNN	93.175	92.037

difficult. The kurtosis is better than RMS, which indicates the fault clearly. However, it increases sharply, which is not easy to track the fault propagation.

The WFE of TB2 bearing run-to-failure data is shown as Fig. 15. Compared with the previous indicators, the WFE shows a good monotonicity, especially in the fault generation stage. The indicator comparison of experimental data is consistent with that of degradation simulation, which further confirms the effectiveness of WFE as a degradation indicator.

5.1.1. Automatic detection of bearing fault occurrence

Vibration signals show only slight difference between normal one before degradation stage and incipient fault. In order to predict the bearing RUL, it is necessary to define the fault occurrence during the running process. In this work, a binary classifier is

constructed to prognosticate fault occurrence automatically. The samples of normal state and severe fault state are selected for training this classifier, while other samples of run-to-failure data for testing. It means that, the classifier can predict the incipient fault automatically according to the information learnt through severe fault samples. The effectiveness and feasibility of this binary classification are verified by cross-validation.

According to vibration signals, samples (500, 1000] of TB1 bearing, samples (500, 1000] of TB2 bearing and samples (300, 450] of TB3 bearing are selected as normal samples, whereas samples (1900, 2100] of TB1 bearing, samples (1700, 2100] of TB2 bearing and samples (700, 900] of TB3 bearing are selected as severe fault samples, which are listed in Table 2 and shown in Fig.16.

Case 1. In case1, only normal and severe fault samples of TB2 and TB3 are used for model training, while all bearing data are

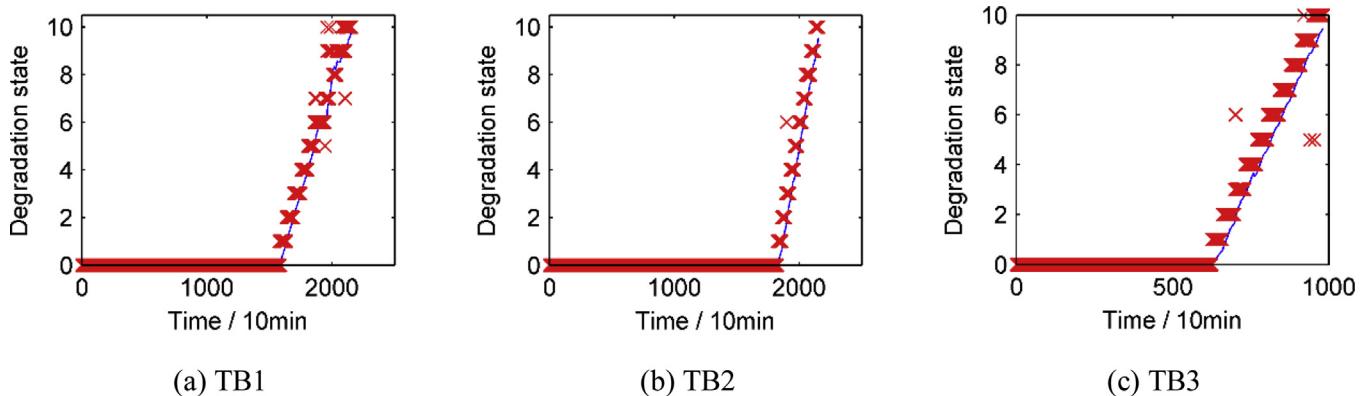


Fig. 19. Identified degradation states of test bearings.

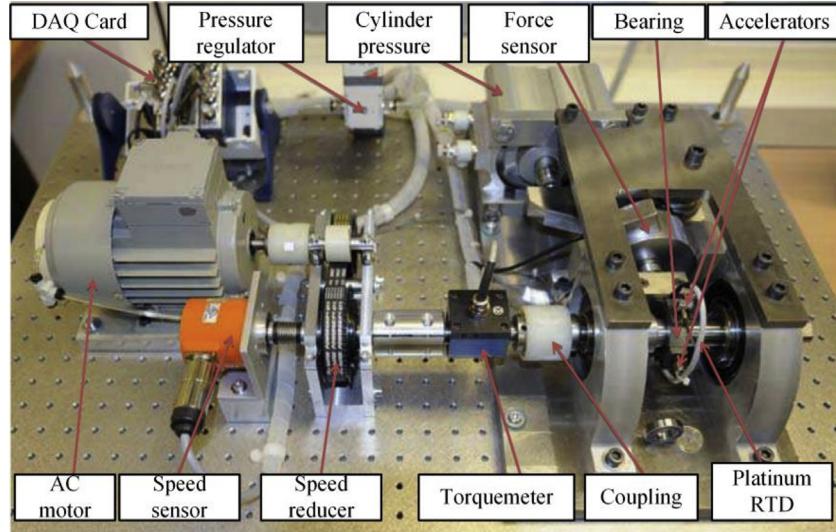


Fig. 20. PRONOSTIA test rig [33].

Table 4

The working condition of each bearing.

Working condition	Rotation / rpm	Load / N	Training bearing	Test bearing
C1	1800	4000	A1 A2	B1 B2 B3 B4 B5
C2	1650	4200	A3 A4	B6 B7 B8 B9 B10
C3	1500	5000	A5 A6	B11

used for testing. Two output categories correspond to the normal and fault possibility of test bearing. Possibility of normality (PON) and possibility of faulty (POF) curve derived from testing are shown in Fig. 17. For TB2 and TB3 bearing data set, the faults are identified in time 1573 and time 603,

respectively. For untrained TB1 bearing, the fault is identified in time 1816 (Fig. 18).

Case 2. In case2, only normal and severe fault samples of TB1 and TB2 are used for model training, while all bearing data are

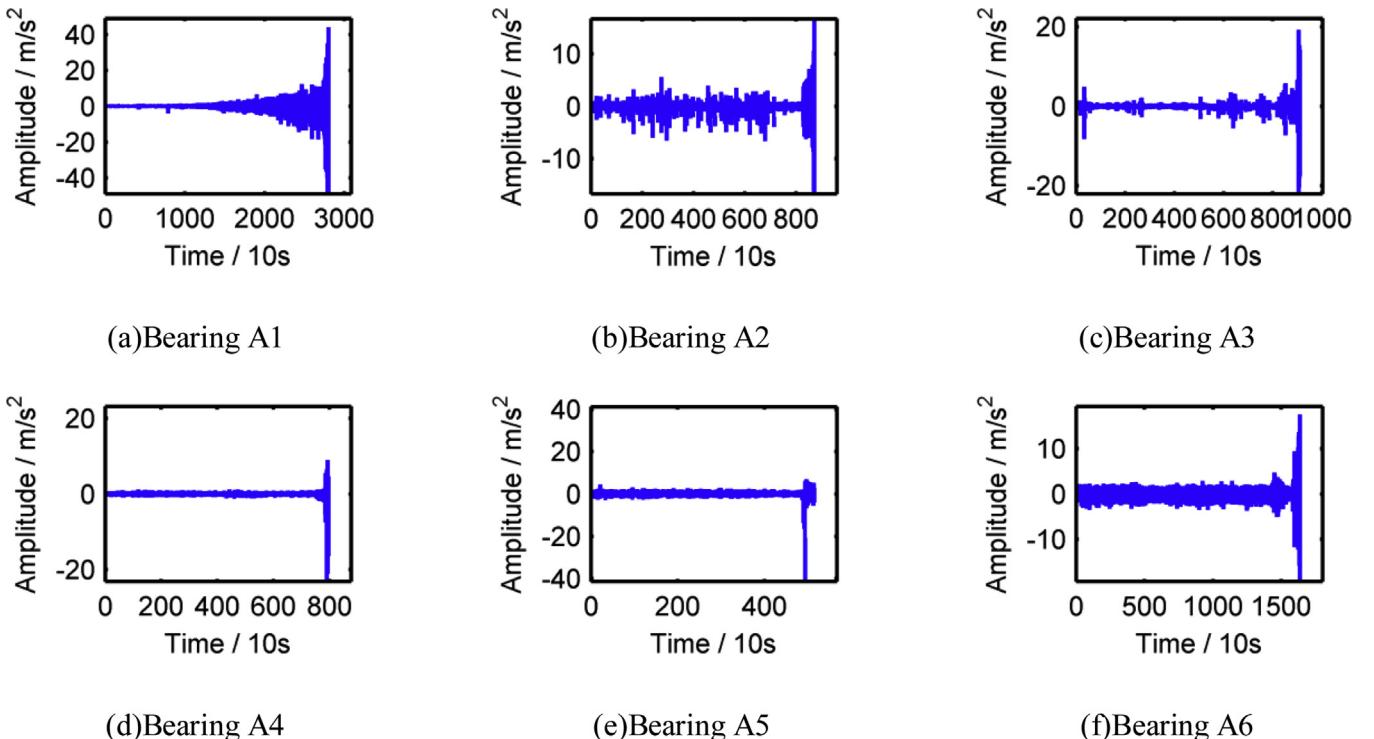


Fig. 21. The degradation process of training bearings.

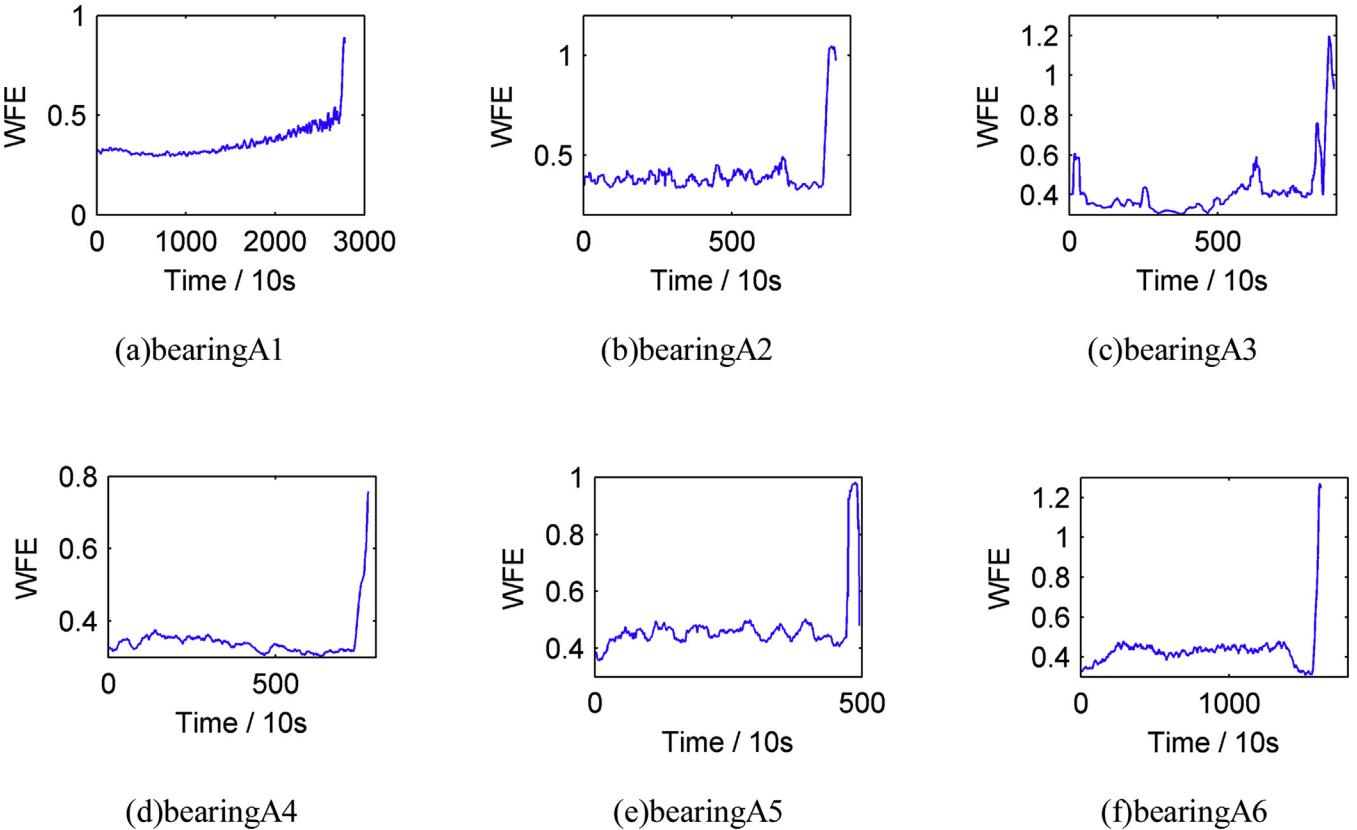


Fig. 22. waveform entropy of training bearings.

used for testing. For bearing TB1 and TB2, the fault occurrences are identified respectively in time 1789 and 1479. For untrained TB3 bearing, the fault is identified in time 619.

Although training with different data sets leads to different detection in these two cases, the experimental results are acceptable. To balance the training difference, we take the mean of two results as the final detection result, in other words, time 1802, 1526 and 611 are identified as fault occurrence point of the three data sets. Compared to FFT or other artificial recognition, this binary classifier needs no experts' experiences and can effectively identify the fault occurrence. For simplicity, the degradation process after fault occurrence is then divided into ten equal stages by time. This binary classification can be regarded as pre-training of final bearing states identification.

5.1.2. Degradation states classification

The LSTM recurrent network is constructed in the TensorFlow and keras framework, and PSO method is used for parameters optimization. Three hidden layers are used and 150 nodes are implemented in each hidden layer. The 'adam' optimizer with a learning rate of 0.001 is applied, and the time step is 8. Totally 5218 samples are obtained after feature extraction and in this section,

70% of the samples are applied for training while other samples are employed for testing. For each fault state, 89 samples are utilized for training while 38 samples for testing.

Back propagation (BP) network, Sparse auto-encoder (SAE) and Convolutional Neural Network (CNN) are employed for comparison. Iterations of all networks are set to 300, and the 'adam' optimizer is also applied for these methods. The structure of BP network is also set to three 150-nodes hidden layers. The structure of SAE is also set to three hidden layers, and the nodes of these three layers are set to 150, 100 and 50, respectively. The input data with 8 time-step is reshape to a vector because SAE and BP cannot process two-dimension data. The CNN applied here have two convolution layers and two pooling layers. The number of convolution kernel are set to 16 and 32, and the size of kernel is set to 5.

Classification accuracy of different methods are shown in Table 3. The combination of kurtosis and waveform factor obtains an accuracy of 78.455%. Compared to previous work [23] (74.2%), it comes to a better accuracy with only traditional features. The classification accuracy markedly increases to 91.748% using waveform entropy and the two traditional features, which illustrates that the combination of both traditional features and entropy indicators can better reflect the bearing degradation level.

Table 5
Fault occurrence identification samples.

Bearing	Failure time / 10s	Normal samples	Severe fault samples	Identified fault occurrence
A1	2803	100-600	2500-2803	2174
A2	871	–	820-871	784
A3	911	200-300	860-911	805
A4	797	50-150	750-797	723
A5	515	100-300	470-515	446
A6	1637	500-600	1600-1637	1521

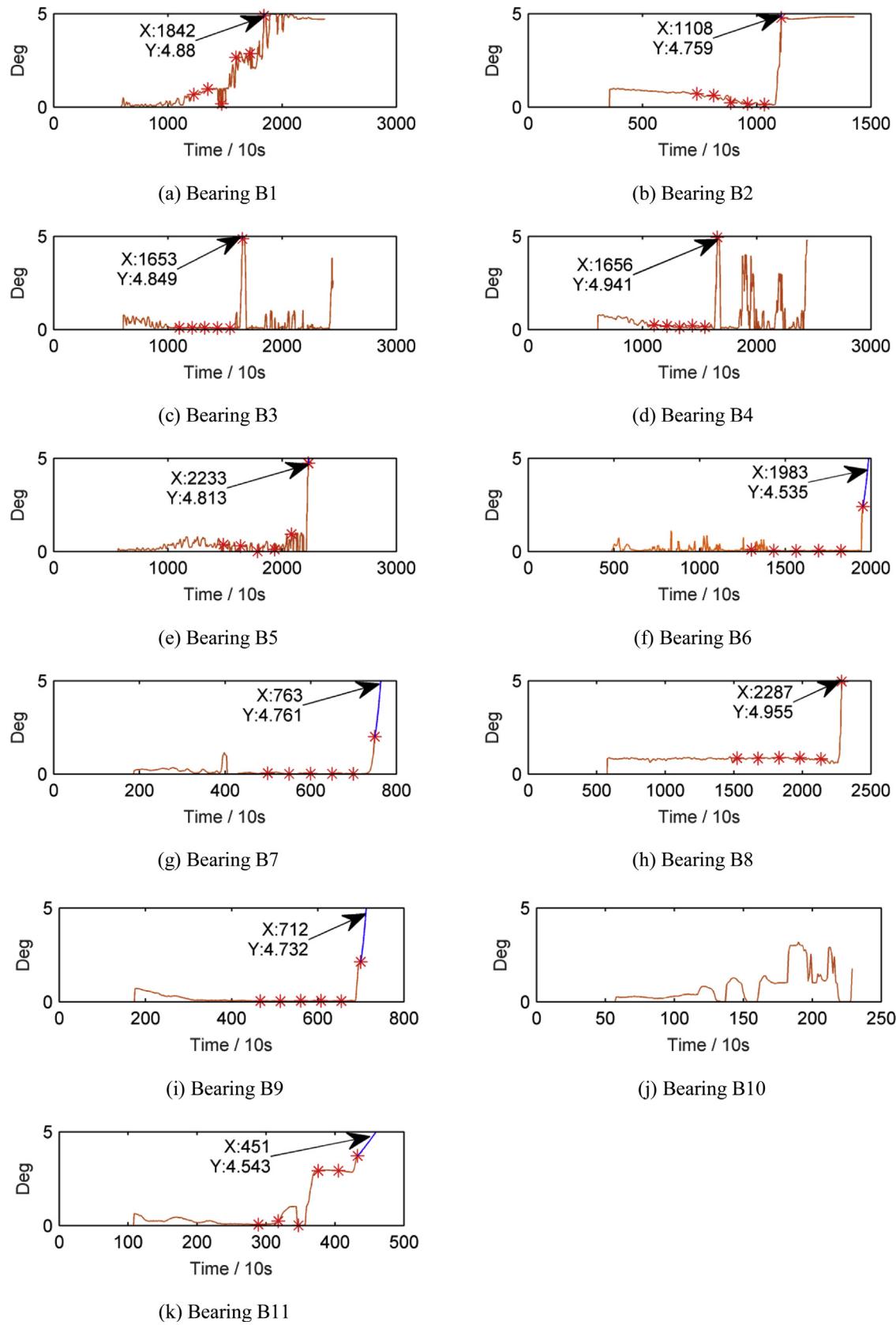


Fig. 23. The degradation state monitoring result of test bearings.

Moreover, waveform entropy impressively brings an accuracy improvement of 14% compared with only traditional features (93.117%), which reveals the excellent degradation assessing ability of waveform entropy. When using back propagation (BP) networks, the training accuracy and testing accuracy were only 81.434% and 78.407%, respectively. SAE had a better performance than BP, but its testing accuracy is no better than 87%. CNN can calculate the local feature of the cover region of its convolution kernel, but its performance is just close to and no better than proposed method. LSTM showed better performance both in training and testing compared with other three networks, which illustrates that LSTM have stronger sequence data processing ability and bearing degradation assessment ability.

The identified bearing degradation states are shown in Fig. 19. As it can be observed, normal state and different fault states can be distinctly recognized by LSTM recurrent network model.

5.2. Bearing health monitoring

The experimental data is obtained from IEEE PHM 2012 Prognostic Challenge, which including a training set A and a testing set B. The vibration signals and temperature signals of 6 whole life bearings are contained in training set while signals of 11 bearings in testing sets. The test rig for bearing degradation experiment is shown in Fig. 20. The bearing shaft is driven by an AC motor, and the bearing pedestal is equipped with a unidirectional vibration acceleration sensor in the vertical and horizontal directions, respectively. In addition, there are holes in the bearing pedestal to install temperature sensors near the outer ring of the bearing. The sampling rate of vibration signals is 25.6 kHz. The vibration data is collected per 10 s, and the sampling length is 0.1 s. The sampling rate of temperature signals is 0.1 Hz. The working condition of each bearing can be seen in Table 4.

Notably, the sampling rate of 25,600 Hz and the sampling length of only 0.1 s make it difficult to analyze the frequency domain indicators.

After denoising by sym4 wavelet, the whole degradation of training bearings is showed in Fig. 21. The amplitude of vibration presents an increasing trend, and the bearing is considered as failure when the amplitude is over 20 m/s².

It can be observed from the time domain signals that bearings under the first working condition have a gently degradation process and a long degradation time. Although wavelet de-noising helpfully removed the interference, the fluctuation still can be observed, especially in bearing A2. As shown in Fig. 22, waveform entropy of all training bearings perform a monotonous trend under all three working conditions. The WFE holds its stationary state until the bearing step into fault stage from normal stage. Even if the vibration signals of the bearing A2 and A3 are affected by noise, the WFE can avoid the saltus and still keep gradual increasing trend, which is beneficial to the degradation prediction.

A binary classification recurrent network is also used to detect the fault occurrence, and the training samples are selected as Table 5. No normal samples are selected from bearing A2 because these samples have big interference.

The identified fault occurrence of A1-A6 bearings are time stamp 2174, 784, 805, 723, 446, and 1521, respectively. According to

the fault occurrence recognition, the degradation process is divided into equal stages by time. Since degradation process of these 6 bearings is relatively short, it is only divided into 5 degradation levels.

After labeling, six kind of bearing states can be obtained. When testing the state of a new bearing, the probability of the six categories of network output is extracted, and the degradation index *Deg* is calculated by the following equation:

$$Deg = \sum_{n=1}^6 p(y=n) * (n - 1) \quad (13)$$

where *n* is the serial number of degradation category, *P* is the possibility of degradation level.

A hard alarm threshold of *Deg* is artificially set to 4.5 because the *Deg* of degradation level is 5. When *Deg* is greater than 4.5, this time is recorded as failure time. However, if the test results are less than 4.5, the following nonlinear exponential degradation model is applied to predict the failure time.

$$y_t = \lambda_1 * \exp(\lambda_2 / t) \quad (14)$$

When applying the model, the zero point and the current *Deg* are considered as the point of fitted curve, and the *Deg* recorded in history would be used to fit the curve by least square method. In order to reduce the computational load, if the last monitoring point is *Tk*, the point with interval of 1/15*Tk between [2/3*Tk, *Tk*] are selected to fit the curve.

The bearing health monitoring results are shown in Fig. 23, and the *Deg* of most test bearings have a monotonous increasing trend. According to the monitoring results, two degradation modes can be identified. The first is slowly growing degradation trend like bearing B1 while another is a rapid one with healing phenomenon like bearing B3 and B4. Considering with the sampling rate of 25.6 kHz and sampling length of 0.1 s, the short signals may not record the fault impact components especially when test bearings have an inner-race fault. And with the development of fault, the bearing performance may enter a repetition stage of "alleviation-aggravation". Hence, the monitored life curve presents an unstable trend in these two bearings (B3 and B4).

The error of predicted failure time and real failure time can be seen in Table 6. The predicted life of the five test bearings (B1-B5) under the first working condition is within the actual life, whereas the lowest error is 1.2% (B5) and the highest is 32.9% (B3). And it can be seen, due to the complexity of B3 and B4 bearings working, it is hard to predict RUL of these bearings accurately. In the second condition, most bearings' RUL are predicted with low error (lower than 1.6%), only the bearing B10 is not recognized effectively because its degradation process is too short to witness the remarkable change of waveform entropy. For bearing B11, working under the third condition, the prediction error is 3.9%.

The experimental results show that the waveform entropy has a good performance under different working conditions, but it has a certain lag due to the window length. For most test bearings, LSTM assessment model has ability to monitor the bearing degradation and can provide guidance for bearing failure warning, but it may not be able to reflect the rapid degradation process completely.

Table 6

The monitoring life error of test bearings.

Bearing	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11
Actual life (10s)	2375	1428	2463	2448	2259	1955	751	2311	701	230	434
Predicted life (10s)	1842	1108	1653	1656	2233	1983	763	2287	712	-	451
Time error (10s)	533	320	810	792	26	28	12	24	11	-	17
Error (%)	22.4	22.4	32.9	32.4	1.2	1.4	1.6	1.0	1.6	-	3.9

6. Conclusion

Bearing performance degradation assessment is crucial to intelligent prognosis, and accurate RUL prediction can provide effective maintenance policy and replacement guidance. In this work, a novel degradation indicator is proposed, and a LSTM recurrent network model is constructed for bearing degradation assessment. Based on the bearing fault vibration response mechanism, the nonlinear degradation signal model is constructed for feature verification and selection. Experimental results demonstrate that the proposed indicator can comprehensively reflect the degradation degree and the LSTM-RNN assessment model can effectively assess the bearing degradation states. However, there are still some issues need to be considered. On the one hand, the exponential degradation component which only reflects the general degradation trend is impossible to simulate the random mutation in the degradation process. On the other hand, the degradation process is simply divided into different stages by time, which might not reflect the real degradation process. These problems will be studied further.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 51475170, 51875208, 51605406).

References

- [1] J. Zarei, J. Poshta, Bearing fault detection using wavelet packet transform of induction motor stator current, *Tribol. Int.* 40 (5) (2007) 763–769.
- [2] L. Cui, J. Huang, F. Zhang, et al., HVS RMS localization formula and localization law: localization diagnosis of a ball bearing outer ring fault[J], *Mech. Syst. Signal Process.* 120 (2019) 608–629.
- [3] Z. Meng, X. Zhan, J. Li, et al., An enhancement denoising autoencoder for rolling bearing fault diagnosis[J], *Measurement* 130 (2018) 448–454.
- [4] L. Song, H. Wang, P. Chen, Step-by-step fuzzy diagnosis method for equipment based on symptom extraction and trivalent logic fuzzy diagnosis theory, *IEEE Trans. Fuzzy Syst.* (2018).
- [5] L. Song, H. Wang, P. Chen, Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery, *IEEE T. Instrum. Meas* 67 (8) (2018) 1887–1899.
- [6] Y. Hao, L. Song, L. Cui, H. Wang, A three-dimensional geometric features-based SCA algorithm for compound faults diagnosis, *Measurement* 134 (2019) 480–491.
- [7] L. Cui, B. Li, J. Ma, Quantitative trend fault diagnosis of a rolling bearing based on Sparsogram and Lempel-Ziv, *Measurement* 128 (2018) 410–418.
- [8] X. Yang, K. Ding, G. He, Double-dictionary signal decomposition method based on split augmented Lagrangian shrinkage algorithm and its application in gearbox hybrid faults diagnosis[J], *J. Sound Vib.* 432 (2018) 484–501.
- [9] L. Cui, Y. Zhang, F. Zhang, J. Zhang, S. Lee, Vibration response mechanism of faulty outer race rolling element bearings for quantitative analysis, *J. Sound Vib.* 364 (2016) 67–76.
- [10] Y. Liao, L. Zhang, W. Li, Regrouping particle swarm optimization based variable neural network for gearbox fault diagnosis, *J. Intell. & Fuzzy Syst* 34 (6) (2018) 3671–3680.
- [11] L. Liao, Discovering prognostic features using genetic programming in remaining useful life prediction, *IEEE Trans. Ind. Electron.* 61 (5) (2013) 2464–2472.
- [12] X. Chen, Z. Shen, Z. He, C. Sun, Z. Liu, Remaining life prognostics of rolling bearing based on relative features and multivariable support vector machine, *Proc. Inst. Mech. Eng. C.-J. Mech. Eng. Sci.* 227 (12) (2013) 2849–2860.
- [13] L. Selak, P. Butala, A. Sluga, Condition monitoring and fault diagnostics for hydropower plants, *Comput. Ind.* 65 (6) (2014) 924–936.
- [14] G. Li, J. Li, S. Wang, X. Chen, Quantitative evaluation on the performance and feature enhancement of stochastic resonance for bearing fault diagnosis, *Mech. Syst. Signal Process.* 81 (2016) 108–125.
- [15] W. Li, S. Zhang, S. Rakheja, Feature Denoising and Nearest–Farthest Distance Preserving Projection for Machine Fault Diagnosis, *IEEE T. Ind. Inform.* 12 (1) (2016) 393–404.
- [16] Z. Chen, W. Li, Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network, *IEEE T. Instrum. Meas.* 66 (7) (2017) 1693–1702.
- [17] Y. Qian, R. Yan, S. Hu, Bearing degradation evaluation using recurrence quantification analysis and Kalman filter, *IEEE Trans. Instrum. Meas.* 63 (11) (2014) 2599–2610.
- [18] Y. Chen, F. Zhu, J. Lee, Data quality evaluation and improvement for prognostic modeling using visual assessment based data partitioning method, *Comput. Ind.* 64 (3) (2013) 214–225.
- [19] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Remaining useful life estimation based on nonlinear feature reduction and support vector regression, *Eng. Appl. Artif. Intel.* 26 (7) (2013) 1751–1760.
- [20] R. Zhao, D. Wang, R. Yan, et al., Machine health monitoring using local feature-based gated recurrent unit networks, *IEEE Trans. Ind. Electron.* 65 (2) (2018) 1539–1548.
- [21] R. Huang, L. Xi, X. Li, C.R. Liu, H. Qiu, Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, *Mech. Syst. Signal Process.* 21 (1) (2007) 193–207.
- [22] J. Yu, Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models, *Mech. Syst. Signal Process.* 25 (7) (2011) 2573–2588.
- [23] J.B. Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, F. Fnaiech, Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network, *Mech. Syst. Signal Process.* 56 (2015) 150–172.
- [24] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, *Neurocomputing*, 240 (2017) 98–109.
- [25] G. He, K. Ding, H. Lin, Fault feature extraction of rolling element bearings using sparse representation[J], *J. Sound Vib.* 366 (2016) 514–527.
- [26] M.I. Jordan, Serial order A parallel distributed processing approach, *Adv. Psychol.* 121 (1997) 471–495.
- [27] J.L. Elman, Finding structure in time, *Cogn. Sci.* 14 (2) (1990) 179–211.
- [28] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [29] M. Liwicki, A. Graves, H. Bunke, et al., A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks, *Proc. 9th Int. Conf. on Document Analysis and Recognition (ICDAR)* 1 (2007) 367–371.
- [30] O. Sutskever, Q.V. Le, Vinyals, Sequence to sequence learning with neural networks, *Advances in neural information processing systems (NIPS)*. (2014) 3104–3112.
- [31] J. Kennedy, R. Eberhart, Particle swarm optimization. icnn'95 – international conference on neural networks, *IEEE* 4 (2002) 1942–1948.
- [32] H. Qiu, J. Lee, J. Lin, G. Yu, Robust performance degradation assessment methods for enhanced rolling element bearing prognostics, *Int. J. Adv. Sci. Eng. Inf. Technol.* 17.3 (2003) 127–140.
- [33] P. Nectoux, R. Gouriveau, K. Medjaher, et al., PRONOSTIA an experimental platform for bearings accelerated degradation tests, *IEEE International Conference on Prognostics and Health Management*. (2012) 1–8.