A Regularized LSTM Method for Predicting Remaining Useful Life of Rolling Bearings

Zhao-Hua Liu[1]   Xu-Dong Meng[1]   Hua-Liang Wei[2]   Liang Chen[1]
Bi-Liang Lu[1]   Zhen-Heng Wang[1]   Lei Chen[1]

[1]School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

[2]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK

Abstract: Rotating machinery is important to industrial production. Any failure of rotating machinery, especially the failure of rolling bearings, can lead to equipment shutdown and even more serious incidents. Therefore, accurate residual life prediction plays a crucial role in guaranteeing machine operation safety and reliability and reducing maintenance cost. In order to increase the forecasting precision of the remaining useful life (RUL) of the rolling bearing, an advanced approach combining elastic net with long short-time memory network (LSTM) is proposed, and the new approach is referred to as E-LSTM. The E-LSTM algorithm consists of an elastic mesh and LSTM, taking temporal-spatial correlation into consideration to forecast the RUL through the LSTM. To solve the over-fitting problem of the LSTM neural network during the training process, the elastic net based regularization term is introduced to the LSTM structure. In this way, the change of the output can be well characterized to express the bearing degradation mode. Experimental results from the real-world data demonstrate that the proposed E-LSTM method can obtain higher stability and relevant values that are useful for the RUL forecasting of bearing. Furthermore, these results also indicate that E-LSTM can achieve better performance.

Keywords: Deep learning, fault diagnosis, fault prognosis, long and short time memory network (LSTM), rolling bearing, rotating machinery, regularization, remaining useful life prediction (RUL), recurrent neural network (RNN).

1 Introduction

Rotating machinery has been widely used in electric power, machinery, aviation, metallurgy, and some military industries. Rolling bearings are one of the most important components in rotating machinery. It has a number of advantages such as high efficiency, low friction, and convenient assembly. However, due to the extremely harsh operating environment, the rolling bearing is also one of the high-risk sub-systems[1]. A literature review shows that many rotating machinery faults are caused by rolling bearing damage[2]. The consequences of rolling bearing failures include the reduction or loss of some system functions. Therefore, the diagnosis and prognosis of rolling bearing faults have become particularly urgent. As a key component of bearing prediction, the remaining useful life (RUL) of the running bearing has drawn increasing attention recently.

There are two popular categories of RUL prediction methods: model-based approaches and data-driven approaches[3]. Model-based methods typically describe mechanical degradation processes by establishing mathematical or physical models and using measurement data to update model parameters[4]. These models include the Gaussian mixture model[5], Markov process model[6], Wiener process model[7], etc. Since the model-based approaches are the combination of expert knowledge and mechanical real-time information, the performance can be improved in terms of the RUL prediction for the bearings.

However, there are also some drawbacks for model-based approaches. For example, these methods can be successfully applied to electronic components and small circuits, but they have limited application to electronic products or systems with complex structure, especially wind turbine systems[8]. Moreover, due to the uncertain measurement such as noise, it is difficult to achieve a model-realistic match for accurate mathematical description of real wind turbines[9]. The identification of model parameters also requires a large amount of experimental and empirical data[10]. These shortcomings may inevitably limit the effectiveness of most model-based methods in practical applications.

However, the data-driven methods based on statistic-

al theory and artificial intelligence theory can overcome shortcomings of the above methods. It uses historical fault data and existing observations to make predictions, and does not rely on physical or engineering principles. With the development of modern signal processing technology and intelligent pattern recognition techniques[11–13], the data-driven fault prognosis method for rolling bearings has been used extensively in industrial applications in recent years[14]. A two-stage bearing life prediction strategy was proposed in [3] by estimating the degradation information and using the enhanced Kalman filter (KF) and the expectation maximization algorithm to estimate the RUL of bearing. In [15], a novel method mixing support vector regression (SVR), support vector machine (SVM), and Hilbert-Huang transform (HHT) was proposed to monitor the ball bearing. Tobon-Mejia et al.[16] proposed a prediction model combining wavelet packet decomposition and mixture of Gaussians hidden Markov model. Singleton et al.[17] presented a forecasting model based on the extended KF, whose parameters were estimated from the extracted features of evolutional bearing faults. In [18], a deep belief network (DBN) based feed-forward neural network (FNN) algorithm was presented to forecast the RUL for the rolling bearing, where DBN was used to extract the features of the vibration signal, and then this FNN algorithm was used for prediction and achieved good results. In [19], an adaptive model was proposed to forecast bearing health, which selected the suitable machine learning method according to the evolution trend of bearing data. Chen et al.[20] proposed a new prediction method by using historical data to build an adaptive neuro-fuzzy reasoning system and establish a time evolution forecasting model of the fault.

With the development of sensor technology, massive data collection in electromechanical equipment becomes available, and data-based methods are utilized for the rolling bearing condition monitoring, which makes the application of artificial neural networks in RUL prediction of rolling bearings receive more and more attention. For example, in [21], the minimum quantization error (MQE) of the self-organizing map (SOM) network was used as a new degradation index. To deal with degraded raw data, the back-propagation neural network and weight application to failure times (WAFT) prediction technique are used to establish the rolling bearing prediction model. In [22], a RUL forecasting approach was presented by utilizing competitive learning, where the statistical properties obtained by using the continuous wavelet transform (CWT) to deal with the data were taken as an input of the recurrent neural network (RNN). The similar defect propagation stages of the monitored bearing are represented by clustering the input data.

The elastic nets can perform grouping in which the factors with strong correlation are often selected or not together. In order to avoid the over-fitting problem, decrease the complexity of the algorithm, and deal with the correlation between features, a label-specific features

learning model combining extreme elastic nets with joint label-density-margin space was presented in [23]. The required label-specific features can be extracted because the sparse weight matrix can be generated by adding the L1 regularization term. In [24], by considering the weighted elastic net penalty and image gradient to solve the super resolution problem, elastic networks were used in constrained sparse representation in face images.

It should be noted that traditional neural networks are composed of shallow learning structures, which may not always sufficiently capture all the most useful information in raw data. With the recent breakthrough of deep learning, RNN can effectively deal with sequence prediction learning problems, such as machine translation, traffic flow prediction and the applications in other fields. However, RNN has a vanishing gradient problem which makes the optimization difficult in some applications. Long short-term memory (LSTM) architecture inherits the traditional advantages in the hidden layer neural nodes of RNN, developing a structure called a memory unit to save history information, and adding three types of gates to control the management of left or reserved historical information, which is valid to capture long-term temporal dependencies. In addition, the hard long time lag problem can be also solved by training LSTM[25]. The new LSTM structure is more robust and applicable than the traditional RNN. Some storage units enable LSTM frameworks to remember a longer period of information and enhance the learning capabilities. Therefore, combining the LSTM network, the RUL prediction of rolling bearings can obtain better performance. In [26], RUL prediction was performed using vanilla LSTM nerves to improve the cognitive ability of the model degradation process, and dynamic differential techniques were used to extract inter-frame information. In [27], a deep learning model based on a one-dimensional convolutional neural network (CNN) and multi-layer LSTM network with attention mechanism was presented to predict the RUL of rotatory machine by extracting the useful features form the original signal. Chen and Han[28] proposed a RUL prediction method based on the LSTM network and principal component analysis (PCA) to predict the trend of health indicator for bearing. LSTM is widely used due to its excellent predictive performance, such as short-term traffic prediction[29], continuous sign language recognition[30], analysis of charge state of lithium batteries[31], and sea surface temperature prediction[32]. In addition, the gated recurrent unit (GRU), as a variant of the LSTM network, is also widely applied in fault prognosis of bearing. For example, Shao et al.[33] proposed a novel prognosis approach based on enhanced deep GRU and complex wavelet packet energy moment entropy to forecast an early fault of the bearing, where GRU was used to capture the nonlinear mapping relationship of the monitoring index defined by complex wavelet packet energy moment entropy and achieved higher prognosis accuracy.

As an important industrial task, precise RUL forecasting of a rolling bearing is still challenging, which mainly includes the following three aspects: 1) There are many factors causing bearing failure such as material deterioration, structure damage, and change of operating environment, which increase the complexity of bearing degradation analysis and greatly hinder the development of RUL prediction technology. Because even for the same type of rolling bearings, their useful life is also very different. 2) With the increase of time series, the traditional data-driven methods may have insufficient ability for feature extraction and difficulty characterizing the complex non-linear function mapping relationship, which leads to the lack of accuracy of long-term prediction. 3) Deep learning methods, such as LSTM, still have the problem of over fitting and may fall into a local minimum, thus leading to failure of RUL prediction. For these reasons, a novel LSTM method called E-LSTM to forecast the RUL of rolling bearings is proposed in this paper. The E-LSTM algorithm consists of an elastic net and LSTM, taking temporal-spatial correlation into consideration to deal with bearing degradation through the LSTM which is made up of a large number of memory units. In the E-LSTM framework, the over-fitting problem is solved by utilizing the regularization term based on the elastic net during the training process of the LSTM network. The results demonstrate that the E-LSTM can obtain more accurate correlation values and high stability that are useful for the bearing RUL forecasting.

The major contributions of this paper are listed as follows:

1) To solve the over-fitting problem in the training process of the LSTM model, an improved LSTM algorithm, called E-LSTM, is presented in this paper. Regularized elastic networks and model parameter optimization including regularization hyperparameters are used in this algorithm, and can be used to perform time series prediction.

2) To effectively represent the nonlinear and non-stationary characteristics of the rolling bearing fault data, based on the proposed E-LSTM model, the rolling bearings RUL forecasting algorithm is developed.

2 LSTM model

2.1 Recurrent neural network

RNN[34] is a recursive neural network whose nodes are directionally connected into a ring, exhibiting dynamic time behavior by its internal state. Unlike the feedforward neural network, RNN can deal with time series effectively in a dynamic way based on its internal memory unit, and can learn the latent features of time series. The structure of the RNN and its hidden layer cell structure are shown in Fig. 1. The hidden layer has a self-circulating edge. As depicted by Fig. 1, the output at time t is relevant to the input at time t and the output at time
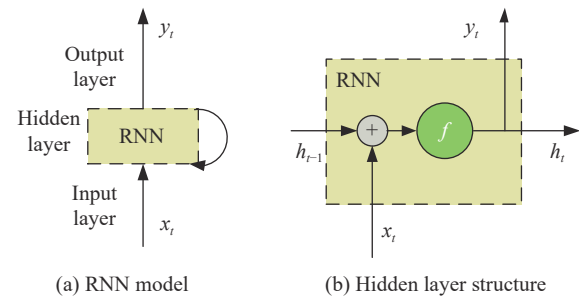

Fig. 1   Structure of the RNN and its hidden layer cell structure. Colored figures are available in the online version.

$t - 1$.

Let the input sequence be $x = (x_1, x_2, \cdots, x_n)$, and $y = (y_1, y_2, \cdots, y_n)$ be the output data. Then, the results of RNN can be described as follows:

$$h_t = f(W_{xt}x_t + W_{ht}h_{t-1} + b_h) \tag{1}$$

$$y_t = W_{hy}h_t + b_y \tag{2}$$

where h_t is the hidden layer state, f denotes the activation function (e.g., tanh function), W represents the matrix in which the weight is replaced (e.g., W_{hy} denotes the weight matrix between hidden layer and output layer), and b represents the bias matrix (e.g., b_h is the bias matrix of hidden layer). The subscript t indicates the time.

Fig. 1(a) shows that the RNN can be viewed as a special case of deep neural networks. When deep neural networks perform the back propagation through time calculation, the deep output error has little effect on the calculation of shallow weights. In other words, the unit of the RNN is mainly affected by the nearby units, meaning that RNN has such a characteristic that its units only have local influence. Therefore, RNN is not capable of dealing with long-term dependencies. As concluded in [35], RNN has the following disadvantages: 1) Due to the gradient vanishing and gradient explosion problem, long delay time series cannot be processed by RNN thoroughly. 2) The predetermined length of the time window is required to train the RNN model. However, it is not easy to automatically get the optimal value of these parameters in the training process.

To overcome these problems, the LSTM model is presented as a special RNN structure. The LSTM model cannot only avoid gradient vanishing, but also learn long-term dependency information.

2.2 LSTM model

The LSTM adopts an improved structure of the original hidden layer neural nodes of RNN, adding a structure called a memory unit to store history information. In addition, input gate, output gate, and forget gate are added in LSTM to determine whether historical information should be removed. As shown in Fig. 2, the hidden

layer cell architecture is more complex than RNN. This LSTM network consists of input gate, output gate, forget gate, and cell state. The input gate controls how much new data can be added to the cell state, the output gate controls the output data of the cell, the forget gate controls the information that should be saved by the cell state, and the cell state is adopted to hold useful information. The forward propagation process of LSTM is expressed as

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \qquad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \qquad (4)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \qquad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \qquad (6)$$

$$h_t = o_t \tanh(c_t) \qquad (7)$$

where i, h, o, f and c are input gate, cell state, output gate, forget gate, and output of the previous cell, respectively. W and b are the weight matrix and bias vector in corresponding units, respectively. σ and tanh are sigmoid and hyperbolic tangent activation functions, respectively.

The LSTM network utilizes the classic back-propagation algorithm to find the optimal parameters during the training, which can be expressed as follows:

1) Based on the forward calculation algorithm, the cell output value $\bar{y}_t$ of LSTM can be calculated as

$$\bar{y}_t = \sigma(\omega_{yh}h_c + b_y) \qquad (8)$$

where $\bar{y}_t$ is the network prediction value at time t, h_c is the state output value of the hidden unit, ω_{yh} is the output weight, and b_y is the output layer bias vector.

2) Reverse calculation of the error term of each LSTM cell. The mean square error of the network prediction is as follows:

$$E_t = \frac{1}{m}\sum_{i=1}^{m}(y_{ti} - \bar{y}_{ti})^2 \qquad (9)$$

where y_{ti} is the i-th true value from the real dataset at time t, and $\bar{y}_{ti}$ is the i-th output value of the LSTM network at time t. m is the number of cells in the output layer of this model. The cumulative error of the model can be obtained from (9) as

$$E = \frac{1}{T}\sum_{t=1}^{T} E_t. \qquad (10)$$

3) Based on the above error obtained, the gradient of all the weights can be calculated. Then the weights will be updated by using the gradient optimization algorithm.
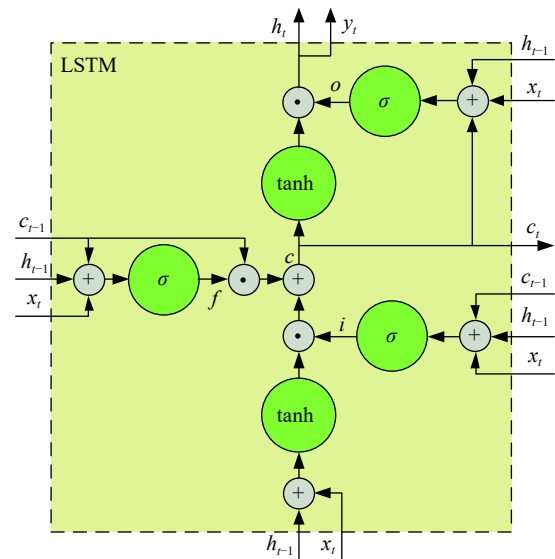
Fig. 2   Hidden layer cell architecture of LSTM

As shown in Fig. 2, it is obvious that the LSTM uses memory cells whose natural behavior is long-term preservation input. To copy the real value of the state and the accumulated external signals, the memory cell in the hidden node can connect weights to itself in the next time step. In addition, the forget gate can be used to determine when the memory contents are cleared. This structure makes it possible for LSTM to predict time series that have long-term dependencies.

3 Proposed E-LSTM network for predicting RUL of rolling bearings

The experimental data collected from traditional rotating machinery are usually non-stationary and noisy[36]. Meanwhile, the traditional LSTM model has an over-fitting problem due to the structural characteristics. Complex working conditions, noise, and over-fitting problems can all make it difficult to carry out accurate prediction. In this paper, an improved regularized LSTM network, called E-LSTM, is proposed to solve the RUL forecasting problem of rolling bearings, and improve its prediction accuracy. The proposed E-LSTM algorithm can not only readily learn the long-term dependence of the process data, but also overcome the over-fitting problem of LSTM for time series prediction.

3.1 Elastic net based model regularization algorithm

The elastic net[37] is the combination of Lasso regularization[34] and ridge regularization[38]. Although the lasso regularization can usually work well for data without strong correlation between features or variables, it is suitable for data modeling problems if there is a high correlations between some features. Ridge regularization can help reduce the variance of the fitted model, while Lasso

regularization can help shrink model coefficients to result in a sparse model, as shown in Fig. 3.
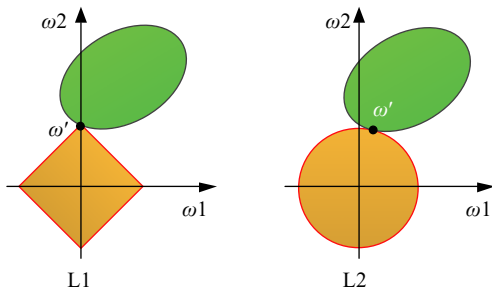

Fig. 3   L1 regularization and L2 regularization

From Fig. 3, it can be seen that the principle of the elastic network is very intuitive. The left side is L1 regularization, and the right side is L2 regularization. The green is the area where the loss function is minimized, and the yellow is the regularization limit area. For L1 regularization and L2 regularization, the optimization goal is to find the intersection of the green area and the yellow area to satisfy the minimization condition of loss function and the regularization limit condition. For L1 regularization, the defined area is a square, and the probability that the intersection of the square and the yellow area is a vertex is very high. There must be $\omega1$ or $\omega2$ at the bump. Therefore, the L1 regularized solution is sparse, which leads to the model preferring to select useful features. For L2 regularization, the defined area is a circle, so that the resulting solution $\omega1$ or $\omega2$ is primarily non-zero and very close to zero. According to the Occam razor principle, a smaller weight means that the network is less complex and the data fits better, thus it can effectively avoid over-fitting problem. By combining the two, the elastic net not only avoids the over-fitting problem but also has stronger feature extraction capability.

The elastic net combines the two regularization methods to achieve complementary effects. After selecting important features, those features that have little or no effect on the life curve will be discarded. The expression of regularization approach is given as follows:

$$\min\left\{\sum_{t=1}^{T} l(y_t, f(u_t, \omega)) + \sum_{i=1}^{m} \lambda_i \rho_i(\omega)\right\} \quad (11)$$

where $l(\cdot, \cdot)$ represents the loss function, which can measure the forecasting performance of the proposed method over the training data set. ω is the model parameters to be estimated, and $\rho(\omega)$ is a regular term used to reduce or avoid over fitting, thus improving the generalization ability of the proposed method. λ is an adjustable regularization parameter. The relationship between the regular term and the loss function is balanced by changing the value of λ.

In this paper, the LSTM network combines the elastic net, and its generalization is enhanced by regularizing the initializing weight ω in the network. The regularization model is expressed as follows:

$$\min\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{m}(y_{ti} - \bar{y}_{ti})^2 + \lambda1||\omega||_1 + \lambda2||\omega||_2^2\right\}. \quad (12)$$

Four different combinations could be obtained by modifying the regularization hyperparameters $\lambda1$ and $\lambda2$ in (12). When $\lambda1 = 0$ and $\lambda2 = 0$, it is a normal LSTM model; when $\lambda1 \neq 0$ and $\lambda2 = 0$, it is the L1 regularization network; when $\lambda1 = 0$ and $\lambda2 \neq 0$, it is the L2 regularization network; when both $\lambda1$ and $\lambda2$ are not equal to 0, it is an elastic regularization network. Following [39], this study employs the combination of L1 and L2 to facilitate important feature selection for LSTM.

The proposed E-LSTM optimization algorithm is utilized to preform RUL forecasting of rolling bearing, and this network structure is illustrated in Fig. 4, where H_{n-1} and C_{n-1} represent the output and cell state of the $(n$-1)-th hidden layer node in the LSTM network respectively, and n is the number of hidden layer nodes in the LSTM network. The representative features of original vibration signals, such as root means square (RMS) value, are extracted and split into training and test samples following the length of the segmentation window as the input of LSTM network. $(x_1, x_2, \cdots, x_i)$ is a input sample and i is the length of the segmentation window and the number of the input nodes in the LSTM network.

$(P_1, P_2, \cdots, P_j)$ represents the predicted outputs of the LSTM network corresponding to $(x_1, x_2, \cdots, x_i)$, and j is the number of the output nodes in the LSTM network. In this study, the number of the output nodes is set to 1. The E-LSTM block diagram consists of the following five parts: input layer, hidden layer, output layer, network optimization, and final prediction. The input layer is in charge of the split and reorganization of the original data to satisfy the input dimensions of the network. The LSTM cell unit shown in Fig. 2 is used to construct
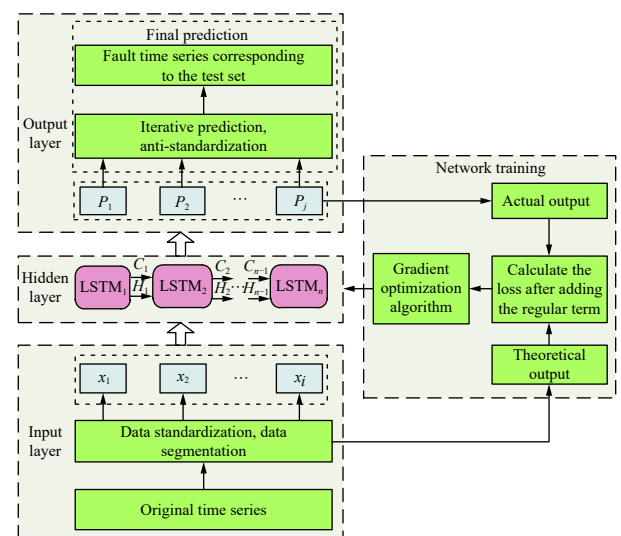

Fig. 4   Training algorithm of E-LSTM model for RUL prediction of rolling bearings

the single hidden layer, and the output layer outputs the predicted values. The elastic net algorithm combining with LSTM network is adopted to train the network, and then a grid optimization algorithm is used to find the optimal regular term hyperparameters. Finally, the stepwise prediction is performed by using the iterative approach.

3.2 Training algorithm

The LSTM neural network is prone to over fitting in the training process, while the elastic net regularization algorithm can shrink the weight of the network by minimizing the loss function. Therefore, optimized by the elastic net regularization algorithm, the LSTM model can overcome the shortcomings of the whole network. Fig. 4 illustrates the training algorithm of the proposed E-LSTM model to forecast the RUL of rolling bearings, and this algorithm is briefly summarized in Algorithm 1.

Algorithm 1. E-LSTM training algorithm

Input: Training data $X_{tr} = \{x_1, x_2, \cdots, x_n\}$ and test data $X_{te} = \{x_{n+1}, x_{n+2}, \cdots, x_m\}$ from the feature extracted from original vibration signal.

Output: The predicted RUL.

1) Randomly initialize the E-LSTM model;
2)   **for** number of training iterations **do**
3)   **for** number of training data **do**
4)   Calculate the predicted value of training data: $\bar{Y}_{tr} = LSTM(X_{tr})$
5)   Calculate the loss by (12);
6)   Update LSTM parameter by back-propagation algorithm;
7)   **end for**
8)   **end for**
9)   Save the trained model $LSTM^*$;
10)   **for** number of test data **do**
11)   Calculate the predicted value of test data: $\bar{Y}_{te} = LSTM^*(X_{te})$
12)   **end for**
13) **return** predicted result $\bar{Y}_{te}$.

The whole RUL forecasting process is depicted in Fig. 5, which consists of the following two parts: offline network training and online forecasting test. The offline network training process performs elastic net based LSTM training until the metric satisfies the requirement. When the training is completed, it is easy to verify the RUL forecasting performance in the testing data. Online RUL forecasting can then be carried out using new E-LSTM network inputs.

4 Experimental study and analysis

4.1 Data source and setup

To verify the effectiveness of the proposed E-LSTM
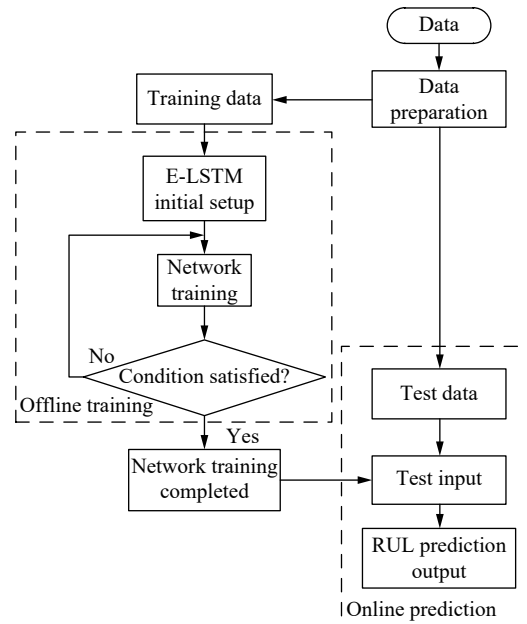
Fig. 5   Schematic description of E-LSTM based rolling bearing RUL prediction process

method, a real-world bearing dataset[40] is used to test in this experiment. These data were collected during the accelerated degradation test of the bearing under different parameters and load conditions through the PRONOS-TIA platform (an experimental platform for bearings accelerated degradation tests). The failure experiments are performed and the experimental data are recorded, as shown in Fig. 6.

Specifically, the motor rotation speed is 1 800 r/min, the load is 4 000 N, the sampling frequency is 25.6 kHz, and the data are recorded every 10 s. There are 7 sets of experimental data in total. Fig. 7 shows the change process of bearing used in the experiment before and after the acceleration test, and Fig. 8 shows the change of the vibration amplitude data collected in a complete accelerated degradation test.

4.2 Feature selection

For predicting the time series, it is essential to select representative features. Commonly used feature values are sometimes combined in the frequency domain, time domain, and time-frequency domain. Different features often represent different physical implications. As reported in [41], the RMS value fairly reflects the overall trend of the rolling bearing data and the abnormal dissipation of the vibration signal energy. Therefore, RMS is used as the experimental feature, which is described as follows:

$$RMS(t) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} X_{ti}{}^2} \tag{13}$$

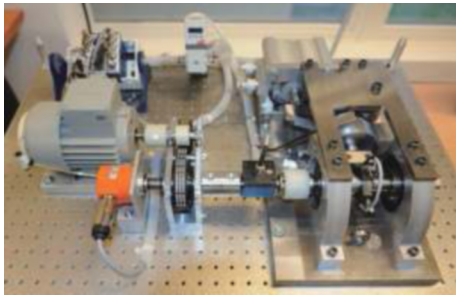where X_{ti} is the i-th original vibration signal at each

Fig. 6   PRONOSTIA platform[40]
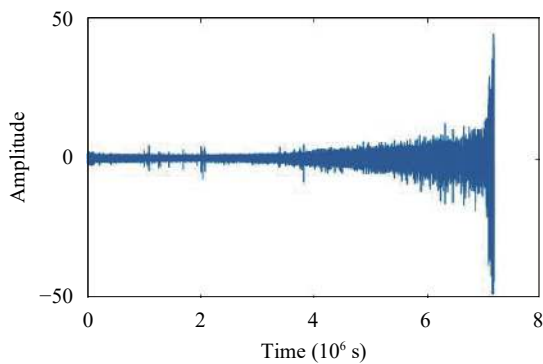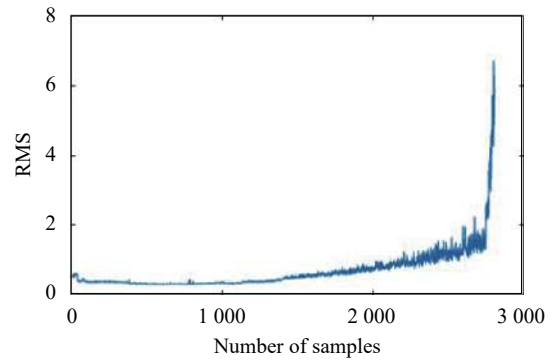

Fig. 7   Normal and degraded bearings[40]

Fig. 8   Original vibration signal curve

(a) Raw RMS curve

(b) Smooth filtered RMS curve
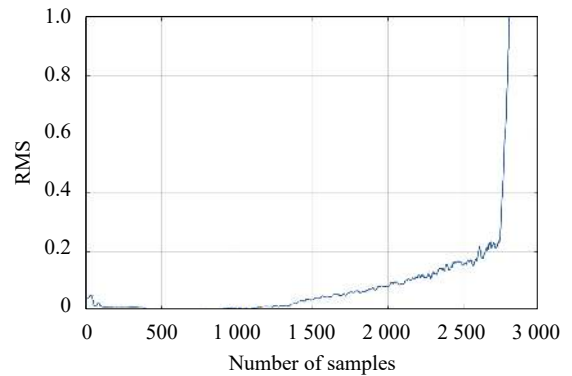
(c) Normalized RMS curve

Fig. 9   Changes of bearing data in the preprocessing process

sampling point t. In addition, N represents the total number of data points collected at the sampling point t, and in this study $N = 2\,560$.
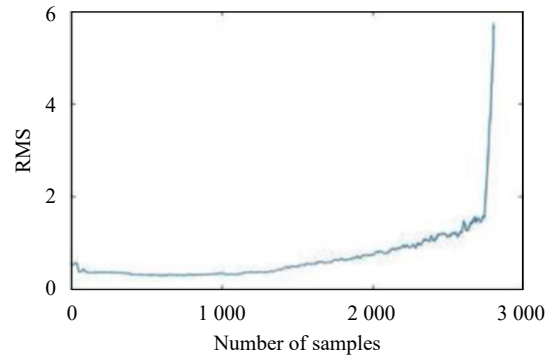
Note that the RMS value is also subjected to mean filtering and normalization under the unified standard to further reduce the noise impact for the RMS signal. The change of rolling bearing data in the whole data processing process is shown in Fig. 9.

4.3   Evaluation of prediction results

The three commonly used metrics for evaluating the performance of time series prediction model are mean square error (MSE), mean relative error (MRE), and mean absolute error (MAE). The MSE metric is more sensitive to the measurement error than the other two[29, 32]. Therefore, MSE is considered as an evaluation criterion for the proposed E-LSTM algorithm. The computing for-mula for MSE is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2 \tag{14}$$

where y_i is the i-th real data, and $\bar{y}_i$ is the i-th predicted data.

4.4   Determination of the LSTM network

The LSTM prediction model involves a large number of parameters. The length of the segmentation window for the model and data should be considered and determined firstly. In order to obtain better prediction performance, the length of data window is investigated in the range of

[1, 10] by trial and error method. The experimental results are shown in Table 1. Fig. 10 shows the MSE value changing as the length of the time window increases. It can be seen that MSE attains its minimum value at 7, meaning that the most acceptable time window length is 7.

The range of the two hyperparameters $\{\lambda 1, \lambda 2\}$ is set to [0, 0.1]. The grid search approach is utilized to find the two optimal hyperparameters in this paper. Compared with other hyperparametric optimization methods (e.g., Bayesian algorithm, genetic algorithm, and particle swarm optimization), the grid search approach is simple, which well meets the experimental requirements of fault diagnosis through time series prediction. For the convenience of calculation, the two hyperparameters are roughly selected from the range of [0, 0.1], and the experimental results are shown in Fig. 11.

From Fig. 11, MSE has an increasing trend with the increase of $\lambda 1$ and $\lambda 2$, but, MSE reaches its minimum (the predefined value obtained by experimental statistical analysis) in the triangle near the zero points (shown in Fig. 12). The regular item parameters are searched iteratively so as to obtain more precise results, and the optimization results are shown in Fig. 12.

From Fig. 12(b), it is known that the MSE value becomes smaller and smaller in the lower right corner region, and thus the optimal values of $\lambda 1$ and $\lambda 2$ are obtained. When $\lambda 1 = 0.009$ and $\lambda 2 = 0.004$, E-LSTM has the best prediction performance. For comparing the prediction accuracy of this proposed model with L1-LSTM (i.e., LSTM with L1 regularization) and L2-LSTM (i.e., LSTM with L2 regularization), it is necessary to find the best performing L1-LSTM method and L2-LSTM method.

Table 1   MSE results of different time window lengths

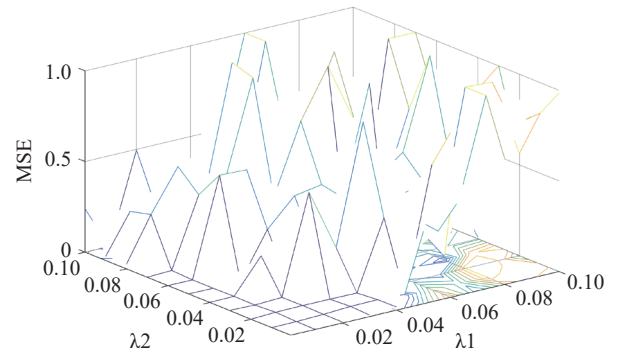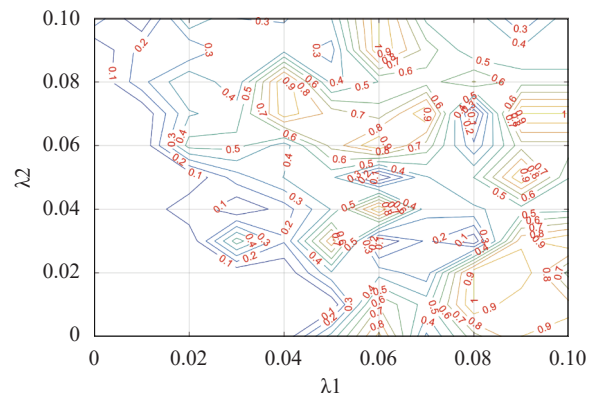| Length	MSE	Length	MSE
1	0.156 17	6	0.077 20
2	0.100 36	**7**	**0.075 29**
3	0.090 15	8	0.083 83
4	0.086 29	9	0.086 40
5	0.077 51	10	0.096 71
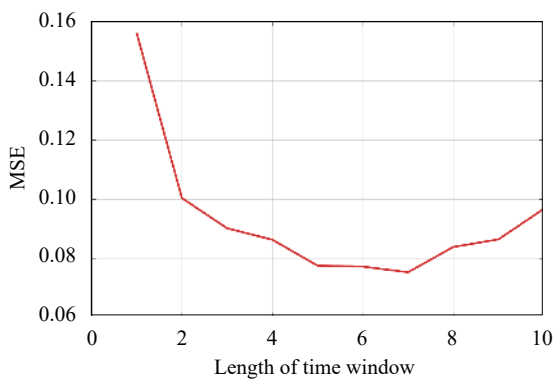
Fig. 10   MSE results of different time window lengths

(a) 3D result diagram of the E-LSTM parameter selection

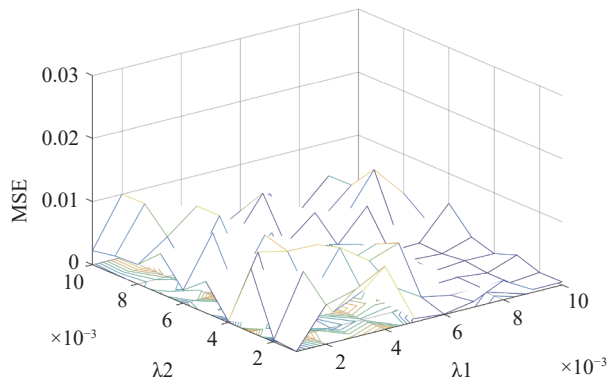(b) Result contour map of the E-LSTM parameter selection

Fig. 11   Parameter rough selection result graph

The hyperparameters of the two models are optimized within a limited range in the experiment, and the results are shown in Fig. 13.

In Fig. 13(a), it is obvious that the MSE value is relatively stable between 0 and 0.02 with the change of $\lambda 1$, but increases rapidly when $\lambda 1 > 0.02$. In order to observe the trend of MSE more accurately, the local amplification of the 0−0.02 range is performed. It is noted that the MSE value decreases first and then increases. Similarly, it is noted from Fig. 13(b) that MSE is stable in the range of 0−0.05. However, the subsequent increase in the MSE value is more stable than that in Fig. 13(a). From the analysis of experimental results, it is concluded that when $\lambda_{L1} = 0.013$, the L1-LSTM model works best, and when $\lambda_{L2} = 0.034$, the L2-LSTM model has the best performance.

4.5   Analysis of experimental results

Through the above experiments for model structure determination and model parameter estimation, three different LSTM models are developed. For making the comparison of the performance of these forecasting methods, i.e., L1-LSTM, L2-LSTM, and E-LSTM, each model is trained and predicted for rolling bearing data. To avoid the influence of accidental factors, 10 independent tests are performed respectively. The statistical values of each group of errors are shown in Table 2.
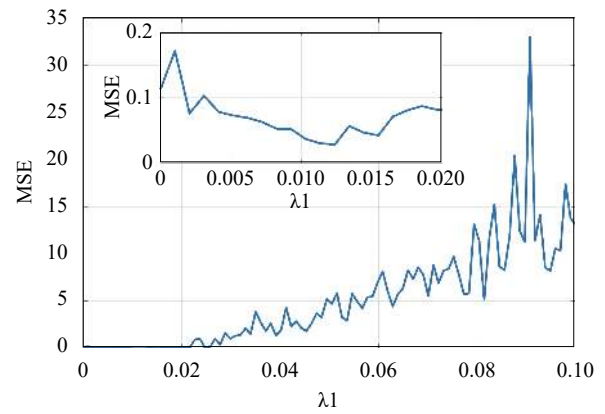
(a) 3D result diagram of the E-LSTM parameter selection
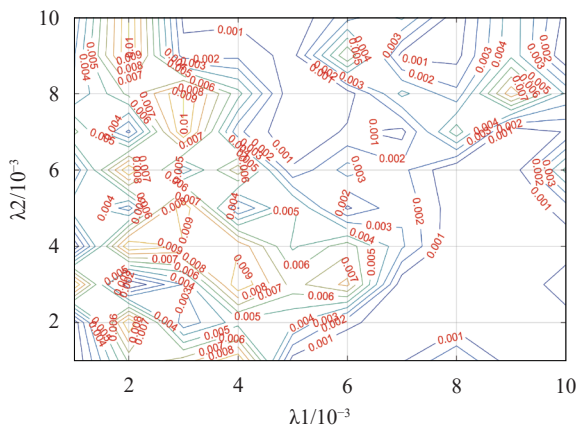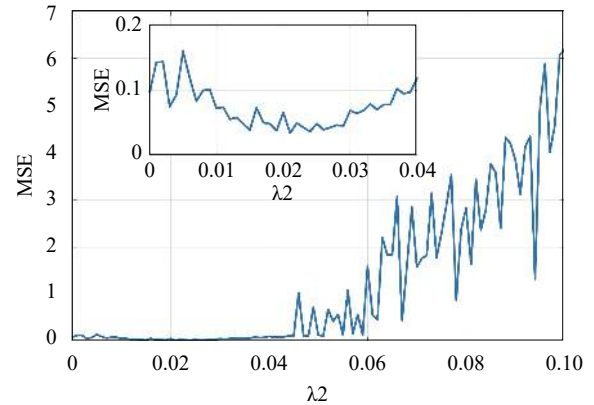

(b) Result contour map of the E-LSTM parameter selection

Fig. 12   Parameter selection resultant graph

(a) L1-LSTM parameter optimization result

(b) L2-LSTM parameter optimization result

Fig. 13   L1-LSTM and L2-LSTM parameter optimization results

Table 2   Comparison of three models with ten tests

Model	MSE value										Mean	Variance
L1-LSTM	0.009 4	0.016 6	0.111 3	0.095 4	0.069 7	0.065 2	0.018 4	0.009 8	0.076 5	0.091 1	0.056 34	1.50×10^{-3}
L2-LSTM	0.060 6	0.107 9	0.076 2	0.047 9	0.076 9	0.010 5	0.083 4	0.050 4	0.155 5	0.030 2	0.069 95	1.70×10^{-3}
E-LSTM	0.029 8	0.048 1	0.031 1	0.018 1	0.028 9	0.019 7	0.018 1	0.009 9	0.018 7	0.016 9	**0.023 93**	**1.17×10^{-4}**

Shown in Table 2, it can be observed that the proposed E-LSTM model outperforms L1-LSTM model and L2-LSTM model in terms of both the mean and variance of the model forecasting errors. For a clearer visualization, the data in Table 2 is presented in Fig. 14.

From Fig. 14, the curve of E-LSTM is not only lower than the other two curves (for most experiments), but also the trend is more stable. It shows that the proposed E-LSTM prediction method can obtain better performance and fairly good robust performance. The algorithm is quite appropriate for RUL forecasting of rolling bearings.

In order to further validate the bearing prediction performance, the comparison is performed between the proposed E-LSTM forecasting algorithm and other five existing approaches, i.e., back propagation neural network (BP), SVM, radial basis function neural network (RBF), DBN, and LSTM network combined with CNN (CNN-LSTM). According to the experimental results, the performance (MSE value) of the six methods is drawn in Fig. 15. It can be seen that the BP and SVM algorithms show roughly the same performance. The performance of the RBF algorithm is slightly better than that of BP and SVM. In addition, deep learning methods (DBN, CNN-LSTM, and E-LSTM) can learn latent features from lots of data and obtain higher prediction accuracy than traditional methods. CNN-LSTM and E-LSTM are the combination of LSTM network and other methods, but the proposed E-LSTM algorithm combines elastic net to avoid over fitting problem in training process and outperforms the CNN-LSTM method.

In order to make detailed comparison, four datasets of bearings obtained in the same work environment (the same speed and loads) are randomly selected, and the prediction is conducted for each case. The datasets are
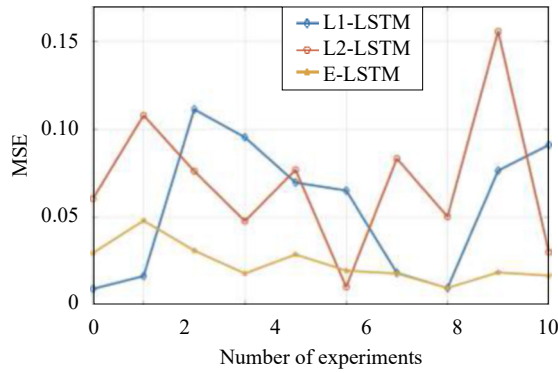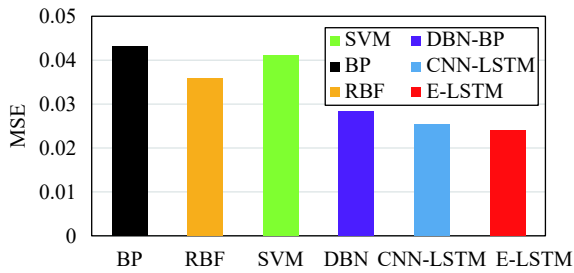
Fig. 14   Comparison of three models with ten tests

Fig. 15   Comparison of mainstream prediction models

denoted as Bearings 1−4. The forecasting results are shown in Fig. 16.

In Fig. 16, the blue curve represents the predicted data, the red curve represents the training data, and the black curve represents the real data. Following [37], in this study, the failure threshold of the bearing data is chosen to be RMS = 0.7 (the solid red line parallel to the X coordinate axis in Fig. 16). Ea represents the intersection abscissa of the actual data curve and the fault threshold line, and Ep represents the intersection abscissa of the predicted data curve and the fault threshold line. The value $Ea - Ep$ describes the discrepancy of the predicted value and the actual value. The value $Ea - Ep$ can be used as an indicator of the model prediction performance. Bearing 4 shows the best predictive performance (Ea and Ep have been overlapped); followed by Bearings 1 and 2, the prediction performance on Bearing 3 is the worst, there is a lag between the true value and the estimated value but the errors of Bearing 3 are not very large. It shows that the E-LSTM algorithm works well for RUL prediction of bearing time series. Meanwhile, the algorithm has good robustness and can forecast the RUL of different bearings in the same work environment.

5   Conclusions

In this paper, an elastic-net regularized LSTM (E-LSTM) method is proposed to forecast the RUL of rolling bearings. The E-LSTM algorithm consists of an elastic net and LSTM, taking temporal-spatial correlation into consideration to deal with the bearing degradation pro-
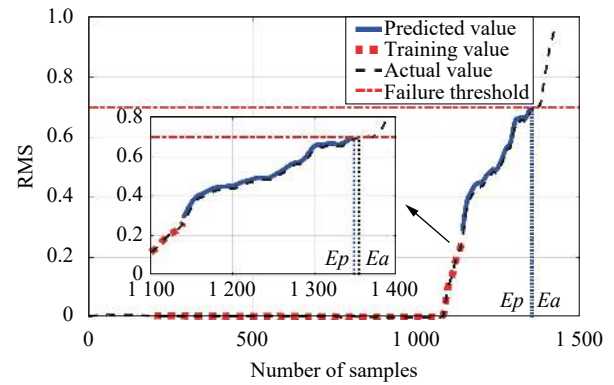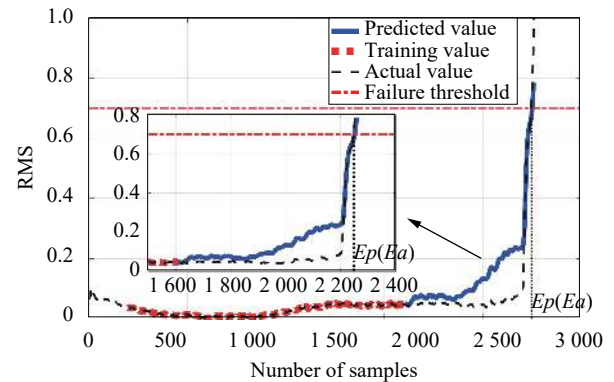

(a) Bearing 1
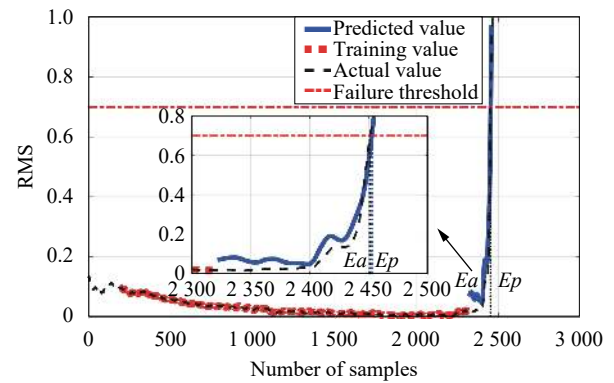
(b) Bearing 2
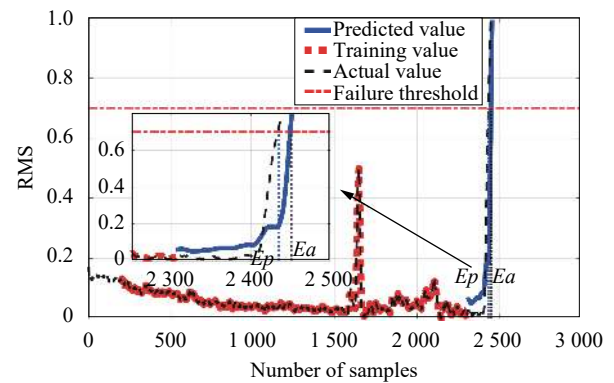
(c) Bearing 3

(d) Bearing 4

Fig. 16   Forecasting results on four bearings test using the proposed method

cess through the LSTM. The elastic net based regularization term is introduced to the LSTM structure to avoid the overfitting problem of the LSTM neural network during the training process. The E-LSTM approach shows better performance than RNN and effectively solves the long-term dependence problem. The combination of the elastic net regularization and the learning ability of LSTM enables the generalization performance of the method proposed which plays an important role in improving the machinery safety of the rolling bearing. However, while the overall forecasting performance of the E-LSTM algorithm is better than the compared methods, the training process of E-LSTM takes more time. So, the future work would be to investigate algorithms to accelerate the calculation speed of E-LSTM and further improve its overall performance for rolling bearing RUL prediction.

Acknowledgements

Open Access

References

[1] H. D. M. de Azevedo, A. M. Araujo, N. Bouchonneau. A review of wind turbine bearing condition monitoring: State of the art and challenges. *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 368–379, 2016. DOI: 10.1016/j.rser.2015.11.032.

[2] B. D. Logan, J. Mathew. Using the correlation dimension for vibration fault diagnosis of rolling element bearings –Ⅱ. Selection of experimental parameters. *Mechanical Systems and Signal Processing*, vol. 10, no. 3, pp. 251–264, 1996. DOI: 10.1006/mssp.1996.0019.

[3] Y. Wang, Y. Z. Peng, Y. Y. Zi, X. H. Jin, K. L. Tsui. A two-stage data-driven-based prognostic approach for bearing degradation problem. *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 924–932, 2016. DOI: 10.1109/TII.2016.2535368.

[4] H. Hanachi, J. Liu, A. Banerjee, Y. Chen, A. Koul. A physics-based modeling approach for performance monitoring in gas turbine engines. *IEEE Transactions on Reliability*, vol. 64, no. 1, pp. 197–205, 2015. DOI: 10.1109/TR.2014.2368872.

[5] J. B. Yu. A nonlinear probabilistic method and contribution analysis for machine condition monitoring. *Mechanical Systems and Signal Processing*, vol. 37, no. 1–2, pp. 293–314, 2013. DOI: 10.1016/j.ymssp.2013.01.010.

[6] H. Y. Dui, S. B. Si, M. J. Zuo, S. D. Sun. Semi-Markov process-based integrated importance measure for multistate systems. *IEEE Transactions on Reliability*, vol. 64, no. 2, pp. 754–765, 2015. DOI: 10.1109/TR.2015.2413031.

[7] X. S. Si, W. B. Wang, C. H. Hu, D. H. Zhou, M. G. Pecht. Remaining useful life estimation based on a nonlinear diffusion degradation process. *IEEE Transactions on Reliability*, vol. 61, no. 1, pp. 50–67, 2012. DOI: 10.1109/TR.2011.2182221.

[8] Y. Q. Cui, J. Y. Shi, Z. L. Wang. Quantum assimilation-based state-of-health assessment and remaining useful life estimation for electronic systems. *IEEE Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2379–2390, 2016. DOI: 10.1109/TIE.2015.2500199.

[9] M. S. Li, D. Yu, Z. M. Chen, K. S. Xiahou, T. Y. Ji, Q. H. Wu. A data-driven residual-based method for fault diagnosis and isolation in wind turbines. *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 895–904, 2019. DOI: 10.1109/TSTE.2018.2853990.

[10] F. Z. Cheng, L. Y. Qu, W. Qiao, L. W. Hao. Enhanced particle filtering for bearing remaining useful life prediction of wind turbine drivetrain gearboxes. *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4738–4748, 2019. DOI: 10.1109/TIE.2018.2866057.

[11] F. Menacer, A. Kadr, Z. Dibi. Modeling of a smart Nano force sensor using finite elements and neural networks. *International Journal of Automation and Computing*, vol. 17, no. 2, pp. 279–291, 2020. DOI: 10.1007/s11633-018-1155-6.

[12] C. J. L. Diaz, D. A. Munoz, H. Alvarez. Phenomenological based soft sensor for online estimation of slurry rheological properties. *International Journal of Automation and Computing*, vol. 16, no. 5, pp. 696–706, 2019. DOI: 10.1007/s11633-018-1132-0.

[13] L. Zhao, X. Wang. A deep feature optimization fusion method for extracting bearing degradation features. *IEEE Access*, vol. 6, pp. 19640–19653, 2018. DOI: 10.1109/ACCESS.2018.2824352.

[14] K. Manohar, B. W. Brunton, J. N. Kutz, S. L. Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, vol. 38, no. 3, pp. 63–86, 2018. DOI: 10.1109/MCS.2018.2810460.

[15] A. Soualhi, K. Medjaher, N. Zerhouni. Bearing health monitoring based on Hilbert-Huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 1, pp. 52–62,

2016. DOI: 10.1109/TIM.2014.2330494.

[16] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, G. Tripot. A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models. *IEEE Transactions on Reliability*, vol. 61, no. 2, pp. 491–503, 2012. DOI: 10.1109/TR.2012.2194177.

[17] R. K. Singleton, E. G. Strangas, S. Aviyente. Extended Kalman filtering for remaining-useful-life estimation of bearings. *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1781–1790, 2015. DOI: 10.1109/TIE.2014.2336616.

[18] J. Deutsch, D. He. Using deep learning-based approach to predict remaining useful life of rotating components. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 1, pp. 11–20, 2018. DOI: 10.1109/TSMC.2017.2697842.

[19] W. Ahmad, S. A. Khan, J. M. Kim. A hybrid prognostics technique for rolling element bearings using adaptive predictive models. *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1577–1584, 2018. DOI: 10.1109/TIE.2017.2733487.

[20] C. C. Chen, B. Zhang, G. Vachtsevanos, M. Orchard. Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering. *IEEE Transactions on Industrial Electronics*, vol. 58, no. 9, pp. 4353–4364, 2011. DOI: 10.1109/TIE.2010.2098369.

[21] R. Q. Huang, L. F. Xi, X. L. Li, C. R. Liu, H. Qiu, J. Le. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 193–207, 2007. DOI: 10.1016/j.ymssp.2005.11.008.

[22] A. Malhi, R. Q. Yan, R. X. Gao. Prognosis of defect propagation based on recurrent neural networks. *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 3, pp. 703–711, 2011. DOI: 10.1109/TIM.2010.2078296.

[23] G. S. Pei, Y. B. Wang, Y. S. Cheng, L. L. Zhang. Joint label-density-margin space and extreme elastic net for label-specific features. *IEEE Access*, vol. 7, pp. 112304–112317, 2019. DOI: 10.1109/ACCESS.2019.2934742.

[24] X. B. Pei, T. Dong, Y. Guan. Super-resolution of face images using weighted elastic net constrained sparse representation. *IEEE Access*, vol. 7, pp. 55180–55190, 2019. DOI: 10.1109/ACCESS.2019.2913008.

[25] S. Hochreiter, J. Schmidhuber. LSTM can solve hard long time lag problems. *In Proceedings of the 9th International Conference on Neural Information Processing Systems, Cambridge, USA*, pp. 473–479, 1997.

[26] Y. T. Wu, M. Yuan, S. P. Dong, L. Lin, Y. Q. Liu. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing*, vol. 275, pp. 167–179, 2018. DOI: 10.1016/j.neucom.2017.05.063.

[27] H. Zhang, Q. Zhang, S. Y. Shao, T. L. Niu, X. Y. Yang. Attention-based LSTM network for rotatory machine remaining useful life prediction. *IEEE Access*, vol. 8, pp. 132188–132199, 2020. DOI: 10.1109/ACCESS.2020.3010066.

[28] Y. H. Chen, B. Han. Prediction of bearing degradation trend based on LSTM. In *Proceedings of IEEE Symposium Series on Computational Intelligence, Xiamen, China*, pp. 1035−1040, 2019. DOI: 10.1109/SSCI44817.2019.900 2776.

[29] Z. Zhao, W. H. Chen, X. M. Wu, P. C. Y. Chen, J. M. Liu. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017. DOI: 10.1049/iet-its.2016.0208.

[30] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, B. B. Chaudhuri. A modified LSTM model for continuous sign language recognition using leap motion. *IEEE Sensors Journal*, vol. 19, no. 16, pp. 7056–7063, 2019. DOI: 10.11 09/JSEN.2019.2909837.

[31] E. Chemali, P. J. Kollmeyer, M. Preindl, R. Ahmed, A. Emadi. Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6730–6739, 2018. DOI: 10.1109/TIE.2017.2787586.

[32] Y. T. Yang, J. Y. Dong, X. Sun, E. Lima, Q. Q. Mu, X. H. Wang. A CFCC-LSTM model for sea surface temperature prediction. *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 207–211, 2018. DOI: 10.1109/LGRS.2017.2780843.

[33] H. D. Shao, J. S. Cheng, H. K. Jiang, Y. Yang, Z. T. Wu. Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing. *Knowledge-Based Systems*, vol. 188, Article number 105022, 2020. DOI: 10.1016/j.knosys.2019.105022.

[34] P. J. Angeline, G. M. Saunders, J. B. Pollack. An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 54–65, 1994. DOI: 10.1109/72.265960.

[35] X. L. Ma, Z. M. Tao, Y. H. Wang, H. Y. Yu, Y. P. Wang. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015. DOI: 10.1016/j.trc.2015.03.014.

[36] J. D. Zheng, H. Y. Pan, S. B. Yang, J. S. Cheng. Generalized composite multiscale permutation entropy and Laplacian score based rolling bearing fault diagnosis. *Mechanical Systems and Signal Processing*, vol. 99, pp. 229–243, 2018. DOI: 10.1016/j.ymssp.2017.06.011.

[37] H. Zou, T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005. DOI: 10.1111/j.1467-9868.2005.00503.x.

[38] A. E. Hoerl, R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970. DOI: 10.1080/00401706.1970.10488634.

[39] F. E. Sloukia, R. Bouarfa, H. Medromi, M. Wahbi. Bearings prognostic using Mixture of Gaussians hidden Markov model and support vector machine. *International Journal of Network Security & Its Applications*, vol. 5, no. 3, pp. 85–97, 2013.

[40] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In *Proceedings of IEEE International Conference on Prognostics and Health Management, Denver, USA*, pp. 1−8, 2012.

[41] S. Hong, Z. Zhou, E. Zio, W. B. Wang. An adaptive method for health trend prediction of rotating bearings. *Digital Signal Processing*, vol. 35, pp. 117–123, 2014. DOI: 10.10 16/j.dsp.2014.08.006.
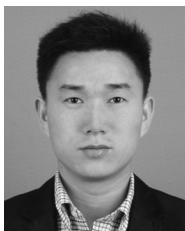
Zhao-Hua Liu received the M. Sc. degree in computer science and engineering, and the Ph. D. degree in automatic control and electrical engineering from Hunan University, China in 2010 and 2012, respectively. He worked as a visiting researcher in Department of Automatic Control and Systems Engineering at University of Sheffield, UK from 2015 to 2016. He is currently an associate professor with School of Information and Electrical Engineering, Hunan University of Science and Technology, China. He has published a monograph in the field of biological immune system inspired hybrid intelligent algorithm and its applications, and published more than 30 research papers in refereed journals and conferences. He is a regular reviewer for several international journals and conferences.

His research interests include artificial intelligence and machine learning algorithm design, parameter estimation and control of permanent-magnet synchronous machine drives, and condition monitoring and fault diagnosis for electric power equipment.

E-mail: zhaohualiu2009@hotmail.com

ORCID iD: 0000-0002-6597-4741

Xu-Dong Meng received the B. Sc. degree in information and communications engineering from Hunan Institute of Technology, China in 2016, and the M. Sc. degree in automatic control and electrical engineering from Hunan University of Science and Technology, China in 2019.

His research interests include machine learning, data mining, and condition monitoring and fault diagnosis for electric power equipment.

E-mail: 1337721766@qq.com

Hua-Liang Wei received the Ph. D. degree in automatic control from University of Sheffield, UK in 2004. He is currently a senior lecturer with Department of Automatic Control and Systems Engineering, University of Sheffield, UK.

His research interests include evolutionary algorithms, identification and modelling for complex nonlinear systems, applications and developments of signal processing, system identification and data modelling to control engineering.

E-mail: w.hualiang@sheffield.ac.uk (Corresponding author)

ORCID iD: 0000-0002-4704-7346

Liang Chen received the B. Eng. degree in automation from Henan University, China in 2018. He is currently a master student in automatic control and electrical engineering, Hunan University of Science and Technology, China.

His research interests include deep learning algorithm design and fault diagnosis of wind turbine transmission chains.
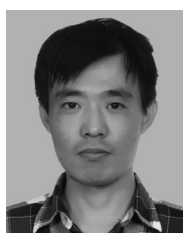
E-mail: 13069302167@163.com

Bi-Liang Lu received the B. Eng. degree in electrical engineering and automation, the M. Sc. degree in automatic control and electrical engineering from Hunan University of Science and Technology, China in 2017 and 2020, respectively.

His research interests include deep learning algorithm design, and condition monitoring and fault diagnosis for electric power equipment.

E-mail: 1197393632@qq.com

Zhen-Heng Wang received the B. Sc. and M. Sc. degrees in automation from Beijing University of Chemical Technology, China in 2006 and 2009, respectively, and the Ph. D. degree in natural resource engineering from Laurentian University, Canada in 2014. Currently, he is a lecturer with Hunan University of Science and Technology, China.

His research interest includes process control, process fault diagnosis and artificial intelligence related subjects.

E-mail: dukehenry83@outlook.com
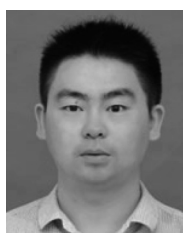
Lei Chen received the M. Sc. degree in computer science and engineering, and the Ph. D. degree in automatic control and electrical engineering from Hunan University, China in 2012 and 2017, respectively. He is currently a lecturer with School of Information and Electrical Engineering, Hunan University of Science and Technology, China.

His research interests include deep learning, network representation learning, information security of industrial control system and big data analysis.

E-mail: chenlei@hnust.edu.cn