Neurocomputing 240 (2017) 98–109

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A recurrent neural network based health indicator for remaining useful life prediction of bearings

Liang Guo[a], Naipeng Li[a], Feng Jia[a], Yaguo Lei[a,b,*], Jing Lin[a]

[a] State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[b] State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu 610031, China

ARTICLE INFO

Article history:
Received 20 December 2016
Revised 8 February 2017
Accepted 15 February 2017
Available online 22 February 2017

Communicated by Hongli Dong

Keywords:
Related-similarity feature
Recurrent neural network
Bearing health indicator

ABSTRACT

In data-driven prognostic methods, prediction accuracy of bearing remaining useful life (RUL) mainly depends on the performance of bearing health indicators, which are usually fused from some statistical features extracted from vibration signals. However, many existing bearing health indicators have the following two shortcomings: (1) many statistical features do not have equal contribution to construction of health indicators since the ranges of these statistical features are different; (2) it is difficult to determine a failure threshold since health indicators of different machines are generally different at a failure time. To overcome these drawbacks, a recurrent neural network based health indicator (RNN-HI) for RUL prediction of bearings is proposed in this paper. Firstly, six related-similarity features are proposed to be combined with eight classical time-frequency features so as to form an original feature set. Then, with monotonicity and correlation metrics, the most sensitive features are selected from the original feature set. Finally, these selected features are fed into a recurrent neural network to construct the RNN-HI. The performance of the RNN-HI is verified by two bearing data sets collected from experiments and an industrial field. The results show that the RNN-HI obtains fairly high monotonicity and correlation values and it is beneficial to bearing RUL prediction. In addition, it is experimentally demonstrated that the proposed RNN-HI is able to achieve better performance than a self organization map based method.

1. Introduction

Rolling element bearings are one of the most critical components in rotating machinery to support rotating shafts. Any unexpected failure of bearings may result in several negative implications, such as downtime increase, productivity reduction and even raise of safety risks [1–5]. To solve these problems, remaining useful life (RUL) prediction is required to schedule a future action to avoid catastrophic events, extend life cycles, etc [6,7]. Therefore, there is an emerging need to develop and improve the techniques of RUL prediction of bearings [8–10].

For RUL prediction of bearings, data-driven methods have been widely investigated [11,12]. As illustrated in Fig. 1, the framework of the data-driven methods is generally based on the following three steps: (1) data acquisition, (2) construction of health indicators (HIs), and (3) prognostics (i.e., prediction of a failure time). Among these steps, the performance of HIs has a crucial effect on prediction accuracy of RUL [13]. In literature, synthesized HIs have received much attention in recent years [14]. The synthesized HIs are usually constructed through some data fusion techniques, where multidimensional statistical features, e.g. root mean square (RMS), variance, kurtosis, etc., are transformed into an one-dimensional HI [15]. Although the synthesized HIs have owned impressive performance and achieved good results for bearing RUL prediction [16,17], there are still two drawbacks which are required to be solved.

(1) Many classical statistical features have different ranges, which cause that these features have no equal contributions to construction of HIs. Specifically, statistical features extracted from a time domain [16,18], a frequency domain [19] and a time-frequency domain [20] of vibration signals exhibit different degradation signatures and different ranges. For example, two features including kurtosis and RMS are plotted in Fig. 2 to show the whole life of a bearing. The kurtosis is sensitive to an early bearing defect and results in a slight degradation trend before a quick bearing degradation at the end of bearing lifetime. The RMS is used to describe the bearing degradation when the casing of the bearing vibrates heavily. It is found that at a failure time the amplitude of the RMS is 4.33, while the amplitude of

* Corresponding author at: State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China
E-mail addresses: yaguolei@mail.xjtu.edu.cn, leiyaguo@gmail.com (Y. Lei).

http://dx.doi.org/10.1016/j.neucom.2017.02.045
0925-2312/

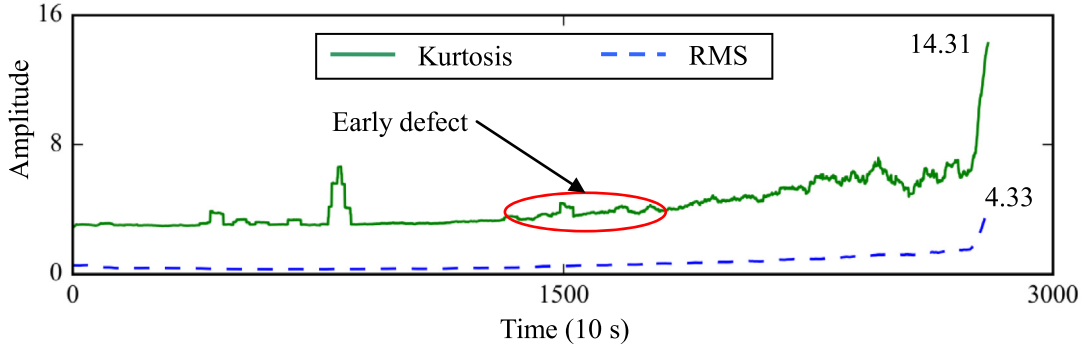


Fig. 1. Data-driven based RUL prediction.

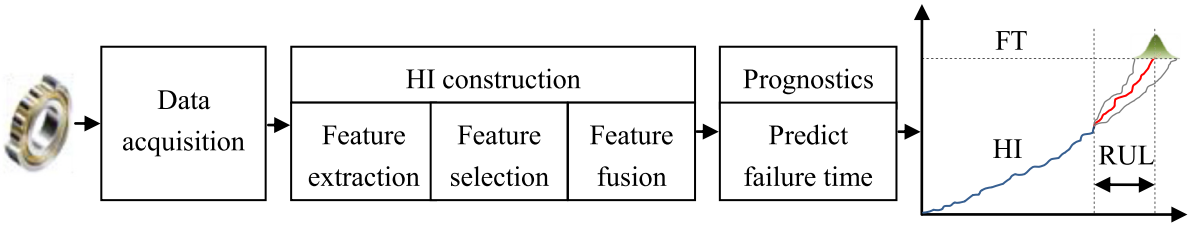


Fig. 2. Two statistical features of a bearing.

the kurtosis is 14.31. Hence, a normalization technique, such as "Min-Max scaling", is usually needed to map statistical features extracted from vibration signals to a specific and equal interval [21,22]. Nevertheless, it is difficult to acquire extreme values of statistical features until some failures occur. This is because the extreme values always appear at the end of bearing lifetime. To address this problem, it is necessary to explore some new techniques which are able to extract statistical features with a specific range. Similarity measurement maps a data sequence to a value ranged from 0 to 1, which may satisfy the aforementioned requirement. Currently, the similarity measurement has been used in some prognostic methods to construct a HI or predict RUL directly. For instance, Medjaher et al. [23] calculated the similarity coefficient of vibration signals between degraded and normal bearings to predict bearing RUL. Yu [24] conducted a HI that was used to calculate the similarity between probability density functions described by two different hidden Markov models and the HI was employed to monitor the performance degradation of bearings. Inspired by the above idea, in this paper, the similarity measurement is utilized to extract features with a specific scale.

(2) It is difficult to determine a failure threshold (FT) since at the failure time HI values of different machines generally have a large variation range. For data-driven methods, RUL is obtained when a HI exceeds a pre-defined FT (as illustrated in Fig. 1). However, the FT is usually determined experimentally, since the HI values of different bearings at a failure time are generally different [13,25]. To address the uncertainty of the FT, Wu et al. [26] proposed a neural network (NN) based RUL prediction method, where its output was the life percentage of a machine. In other words, the output was expected to one (one hundred percent) when the machine was in a failure condition. Compared with the aforementioned NN, recurrent neural network (RNN) is a more effective model that involves time-series data [6,27]. For instance, Tse et al. [28] used RNN to forecast successive HI values at the next time step. Felix [29] proposed a RNN based method to predict RUL of a Turbofan engine. Qian et al. [30] input some extracted features directly to a RNN model to predict RUL of bearings. Although the performance of the RNN for RUL prediction has been investigated in recent years, the RNN based method may be confronted with a problem that prediction results do not provide any confidence limit [6,13]. Therefore, in this paper, the RNN is applied to construct a HI, and the RUL and its associated prediction intervals are calculated through an exponential model.

In order to deal with the aforementioned two shortcomings, this paper presents a RNN based HI for RUL prediction of bearings. First, to overcome the barrier of large variances of classical statistical features, six related-similarity (RS) features are proposed to be combined with eight time-frequency features so as to form an original feature set that contains rich degradation signatures of bearings. Then, the monotonicity and correlation metrics are used to select the most sensitive fault features. Finally, these selected features are fused into a HI (RNN-HI) through a RNN. The performance of the proposed RNN-HI is validated by data sets from experiments and an industrial field. The main contributions of this paper are summarized as follows.

(1) A feature extraction method, i.e. RS, is proposed and used to mine useful degradation information. The RS calculates the similarity between the current inspection data and data at an initial operation point as a feature which ranges from 0 to 1. Therefore, the RS features can be directly used for RUL prediction without normalization.
(2) A RNN is trained to automatically map some selected features to a RNN-HI which is the degradation percentage of a bearing. The value of the RNN-HI at a failure condition equals to one, and therefore a FT does not need to be specified artificially. In addition, the performance of the RNN-HI is verified through the experiment and industrial field data.

The rest of the paper is organized as follows. Section 2 introduces the basic theory of RNN. The detailed construction procedure of the proposed RNN-HI is then presented in Section 3. Using the data set from the accelerated degradation testing on rolling

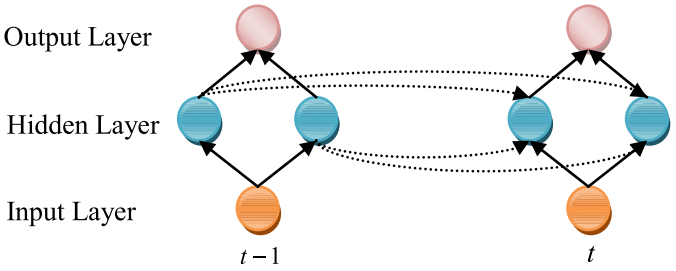


Fig. 3. A recurrent neural network.

element bearings, the performance of the RNN-HI is evaluated in Section 4. Then, in Section 5, it is validated through a data set from an industrial field. Finally, conclusions are drawn in Section 6.

2. Basic theory of RNN

Compared with a feed-forward NN, a RNN is strengthened by a time-step edge that introduces a notion of time to the NN model. And edges connecting adjacent steps, called recurrent edges, form cycles that are self-connections of a neuron to itself across time. As shown in Fig. 3, at time t, recurrent neurons are input with information not just from the previous layer x_t but also from themselves of the previous position h_{t-1}. Consequently, the output $\hat{y}_t$ is influenced not only by the current input information but also by the information at time $t-1$. Mathematically, these processes can be described using the following transition function:

$$h_t = f(w_{hx}x_t + w_{hh}h_{t-1} + b_h) \tag{1}$$

$$\hat{y}_t = f(w_{yh}h_t + b_y) \tag{2}$$

where $f(\cdot)$ is an activation function. w_{hx} is the matrix of conventional weights between an input layer and a hidden layer and w_{hh} is the matrix between a hidden layer and itself at adjacent time steps. The vectors b_h and b_y are bias parameters which allow each node to learn an offset.

As exhibited in Fig. 3, the structure of the RNN across a time can be described as a deep network with one layer per time step. It can be seen that the network can be trained across time steps using backpropagation that is called backpropatation through time [31]. But the learning process is especially challenging due to the problem of gradients vanishing or exploding. This is the main reason why the training optimal process approaches to a local extreme value.

To overcome the problem of gradient vanishing or exploding, a long short-term memory (LSTM) architecture that involves a memory cell was constructed [32]. The memory cell replaces hidden neurons used in traditional RNNs to build a hidden layer. There is one node (an internal state s) in the memory cell that connects itself with a fixed weight. In addition, an input gate, an input node, a forget fate and an output are added into the memory cell as shown in Fig. 4, which relaxes the gradient vanishing or exploding problem. Such structure makes sure that the RNN model has the long short-term memory in the form of weights and ephemeral activations.

Mathematically, the computation of the LSTM model can be described as follows:

$$g_t = \phi(w_{gx}x_t + w_{gh}h_{t-1} + b_g) \tag{3}$$

$$i_t = \sigma(w_{ix}x_t + w_{ih}h_{t-1} + b_i) \tag{4}$$

$$f_t = \sigma(w_{fx}x_t + w_{fh}h_{t-1} + b_f) \tag{5}$$

$$o_t = \sigma(w_{ox}x_t + w_{oh}h_{t-1} + b_o) \tag{6}$$

$$s_t = g_t \otimes i_t + s_{t-1} \otimes f_t \tag{7}$$

$$h_t = \phi(s_t) \otimes o_t \tag{8}$$

where w_{gx}, w_{ix}, w_{fx} and w_{ox} are weight values between an input layer x_t and a hidden layer h_t at time t, respectively; w_{gh}, w_{ih}, w_{fh} and w_{oh} are hidden layer weight values between time t and $t-1$, respectively; b_g, b_i, b_f and b_o are bias of the input node, the input gate, the forget gate and the output gate, respectively. h_{t-1} is the output values from the hidden layer at the previous time. g_t, i_t, f_t and o_t are output values of the input node, the input gate, the forget fate and the output gate, respectively. s_t is an internal state at the current time, and s_{t-1} is an internal state value at the previous time. $\otimes$ is the pointwise multiplication operator.

3. Proposed health indicator

This section describes the procedure for construction of the proposed RNN-HI. As shown in Fig. 5, it is mainly composed of three stages, including feature extraction, selection of sensitive features and construction of the RNN-HI.

3.1. Related-similarity and feature extraction

The RS features are calculated through the similarity measurement of data sequences between the current and initial times. If a data sequence at time t is denoted as f_t and a data sequence at

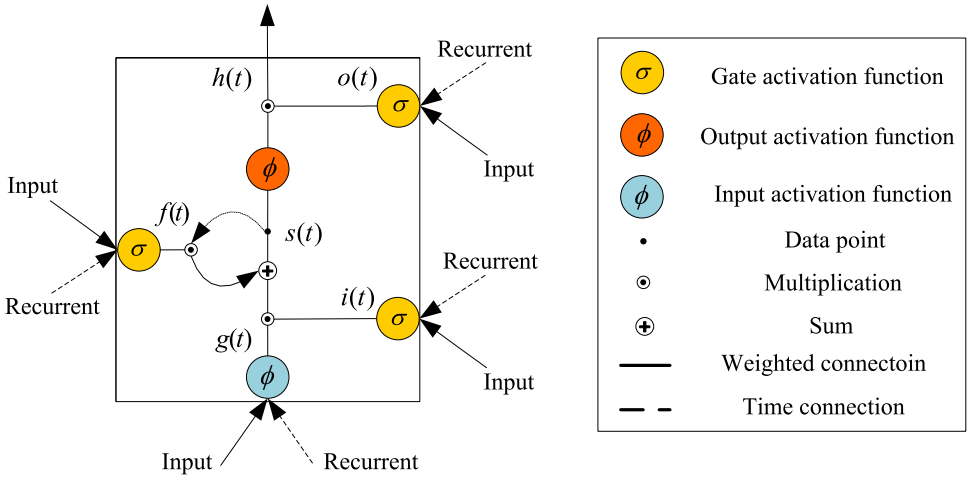


Fig. 4. The architecture of a LSTM memory cell.

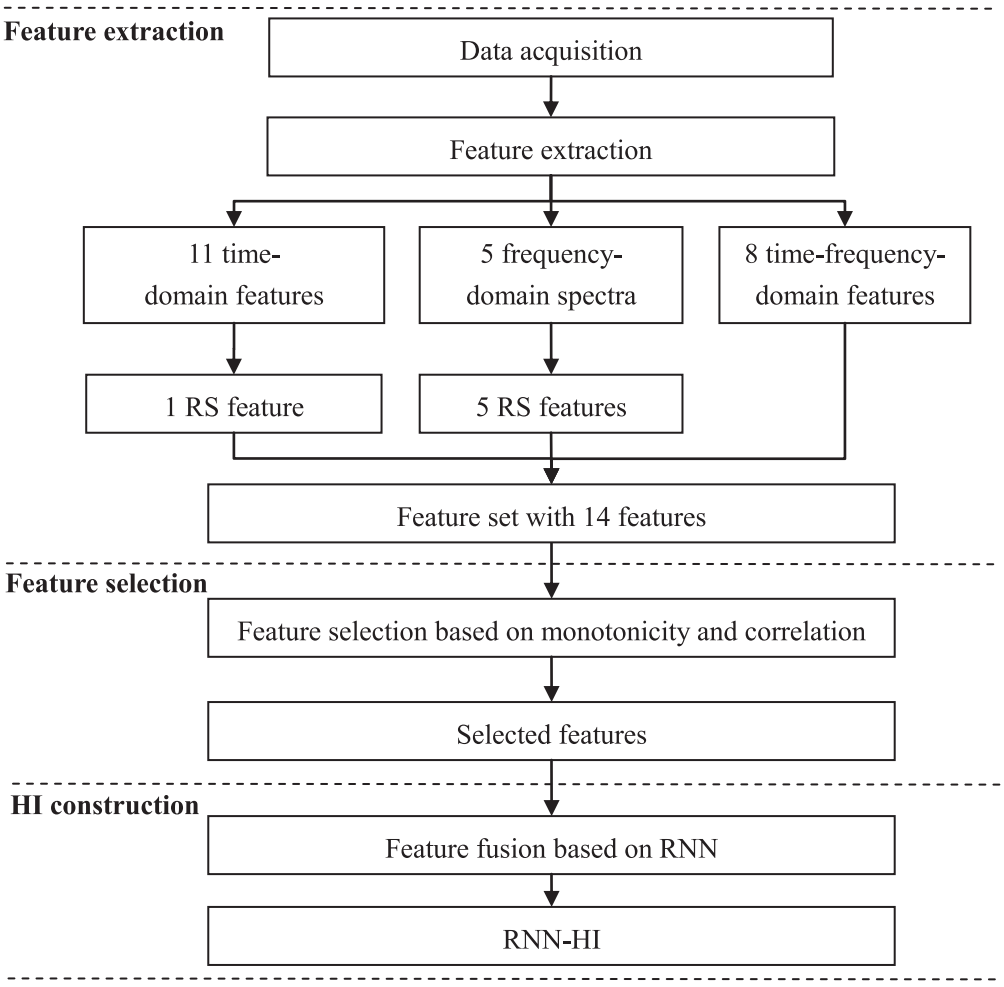


Fig. 5. A flowchart of the proposed health indicator.

an initial operation time is denoted as f_0, the RS features can be calculated by:

$$RS_t = \frac{\left|\sum_{i=1}^{k}(f_0^i - \tilde{f}_0)(f_t^i - \tilde{f}_t)\right|}{\sqrt{\sum_{i=1}^{k}(f_0^i - \tilde{f}_0)^2 \sum_{i=1}^{k}(f_t^i - \tilde{f}_t)^2}} \tag{9}$$

where k is the length of the data sequence. $\tilde{f}_0$ and $\tilde{f}_t$ are the mean value of $\{f_0^i\}_{i=1:k}$ and $\{f_t^i\}_{i=1:k}$, respectively. Fig. 6 illustrates an example of the proposed RS feature in a frequency domain. In Fig. 6, The curves are the frequency spectra of the raw vibration signals in different health conditions, where the green dashed curve is the frequency spectrum f_0 at an initial time of an operating machine and the red dotted curve is the current frequency spectrum f_t. Then, the RS feature RS_t can be obtained through Eq. (9) which is shown as a red star in Fig. 6. Note that if the current frequency spectrum f_t is overlapped extensively with the initial frequency spectrum f_0, the RS feature will be equal to one, which means that the machine stays healthy. Otherwise, it will decrease as machine degrades. Therefore, the RS features are limited to a range from 0 to 1 for different machines.

As shown in Fig. 5, in this paper, the feature set includes one RS feature in a time domain, five RS features in a frequency domain and eight classical features in a time-frequency domain. In the time domain, a data sequence f consists of eleven statistical features, including mean, RMS, kurtosis, skewness, peak-to-peak, variance, entropy, crest factor, wave factor, impulse factor and margin factor. In the frequency domain, a data sequence f is constructed by the frequency spectra, where they are a full frequency spectrum and four sub-bands frequency spectra. Suppose that the sample frequency is 25.6 kHz, and then the full frequency spectrum and four sub-bands frequency spectra are located in 0–12.8 kHz, 0–3.2 kHz, 3.2–6.4 kHz, 6.4–9.6 kHz and 9.6–12.8 kHz, respectively. In addition, eight time-frequency domain features, energy ratios of eight frequency sub-bands generated by performing haar wavelet package transform with a three-level decomposition on vibration signals, are also added into the feature set. The information about these features is detailed in Table 1.

Table 1
Feature sets.

RS features				Energy ratio features	
Time-domain		Frequency-domain		Time-frequency-domain	
*F*1	RS of 11 classical time-domain features	*F*2	RS of [0,12,800]Hz	*F*7	Energy ratio of (3,0)
		*F*3	RS of [0,3200]Hz	*F*8	Energy ratio of (3,1)
		*F*4	RS of [3200,6400]Hz	*F*9	Energy ratio of (3,2)
		*F*5	RS of [6400,9600]Hz	*F*10	Energy ratio of (3,3)
		*F*6	RS of [9600,12,800]Hz	*F*11	Energy ratio of (3,4)
				*F*12	Energy ratio of (3,5)
				*F*13	Energy ratio of (3,6)
				*F*14	Energy ratio of (3,7)

3.2. Selection of sensitive features

The main task of feature selection is to discard irrelevant and redundant features which do not provide sufficient fault

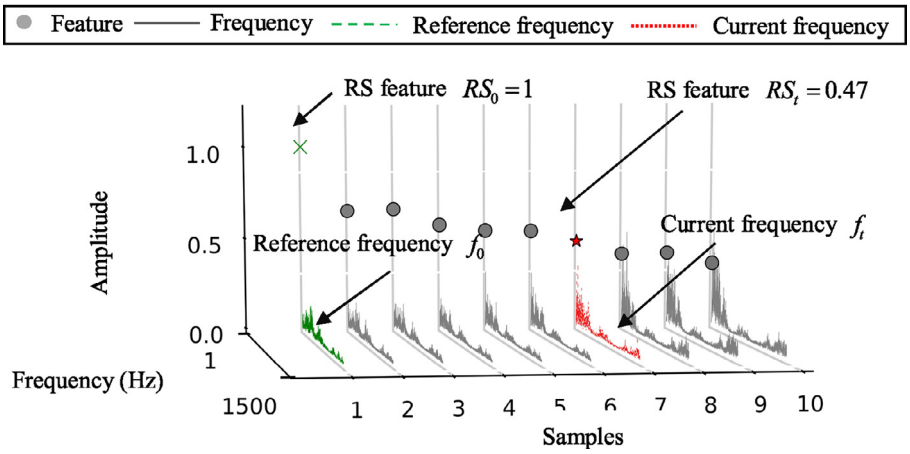


Fig. 6. RS feature in a frequency domain.

signatures. It means that a good prognostic feature should be monotonically correlated with degradation process. Based on these principles, here, the correlation and monotonicity metrics are utilized to select the most sensitive features from the feature set.

The correlation metric measures a linear correlation between features and operating time and it is calculated by [33]:

$$\mathrm{Corr} = \frac{\left|\sum_{t=1}^{T}(F_t - \tilde{F})(l_t - \tilde{l})\right|}{\sqrt{\sum_{t=1}^{T}(F_t - \tilde{F})^2 \sum_{t=1}^{T}(l_t - \tilde{l})^2}} \tag{10}$$

where F_t and l_t are the feature and time values of the tth observation sample. T is the length of the samples during the lifetime.

The monotonicity metric evaluates an increasing or decreasing trend of the features as follows [34]:

$$\mathrm{Mon} = \left|\frac{dF > 0}{T-1} - \frac{dF < 0}{T-1}\right| \tag{11}$$

where dF is the differential of feature series, and T is the length of the samples during the lifetime. Mon = 1 means that the feature is totally monotonic, otherwise it is oscillating.

It can be seen that these two metrics of features are confined in the range [0, 1] and they are positively correlated with the performance of the candidate features, which makes them suitable as the feature selection metrics. Therefore, a linear combination of the two above metrics is calculated as the feature selection criteria:

$$\mathrm{Cri} = \frac{\mathrm{Corr} + \mathrm{Mon}}{2} \tag{12}$$

where Cri is a criteria to select the most sensitive features.

3.3. RNN-HI construction

In order to take advantages of mutual information from these selected features, the RNN is used to fuse the selected features as the RNN-HI. Concretely, the proposed RNN-HI consists of two steps: a training step and a testing step. In the training step, the lifetime samples of bearings are used to form a training set $\{x_t, y_t\}_{t=1}^{T}$, where $x_t \in R^{N \times 1}$ is selected N features at time t, $y_t \in [0, 1]$ is its associated label which indicates the degradation percentage of bearings at time t. For example, suppose that the failure time of a bearing is 2800 s, and the current inspection point is 1400 s, and then the label y_t is 0.5. Therefore, a RNN model is trained through minimizing the following cost function:

$$C = \frac{1}{2}\sum_{t=0}^{T}\left\|y_t - \hat{y}_t\right\|_2^2 \tag{13}$$

where $\hat{y}_t$ is an output of the RNN model, and y_t is a true label value. In this paper, in order to limit the output to a range from 0 to 1, a sigmoid function is chosen as the activation function of output layers. To solve the problem of gradient exploding or vanishing, the LSTM cell is utilized to construct the hidden layer, where the computing process can be found from Eqs. (3)-(8). In the testing step, the selected features of the testing data are directly input to the trained RNN to obtain the RNN-HI that is the proposed HI. The RNN-HI is expected to have a constant FT which equals to one. Therefore, the RUL is able to be estimated depending on the constant FT.

4. Experiment verification

4.1. An experiment system

In this case, the experimental data set from PRONOSTIA was collected by conducting accelerated degradation tests of bearings. Two accelerometers were horizontally and vertically mounted on the bearing to monitor its vibration. Vibration signals collected by the two accelerometers were sampled every 10 s, and the duration of the sampling lasted 0.1 s with a sampling frequency 25.6 kHz. The detailed information about the platform and experiments can be found in [35].

In the experiments, there were three different operating conditions: the first one (1800 rpm and 4000 N), the second one (1650 rpm and 4200 N) and the third one (1500 rpm and 5000 N). The numbers of the testing bearings of these three operation groups were 7, 7 and 3 with named bearing1_1-bearing1_7, bearing2_1-bearing2_7 and bearing3_1-bearing3_3, respectively. Among them, the first two bearings in every group were regarded as a training set and the others were used as a testing set. Fig. 7 shows the vibration signal of a tested bearing during its whole life cycle, it is seen that the amplitudes of vibration signals increase over time, which indicates that vibration signals play a significant role in bearing performance degradation assessment.

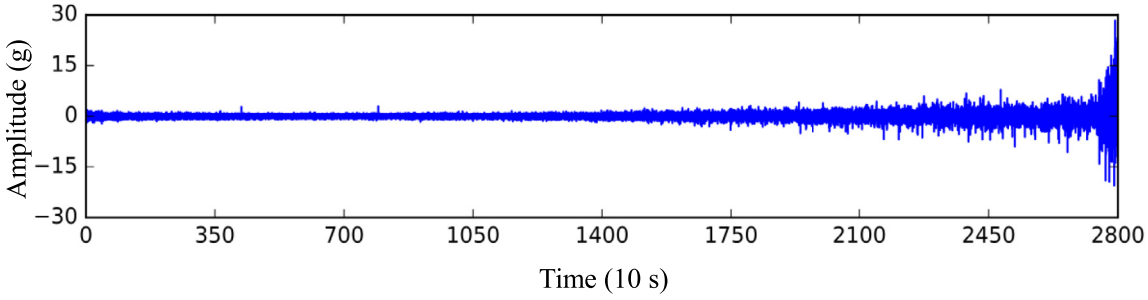


Fig. 7. The temporal vibration signal of bearing1_1.

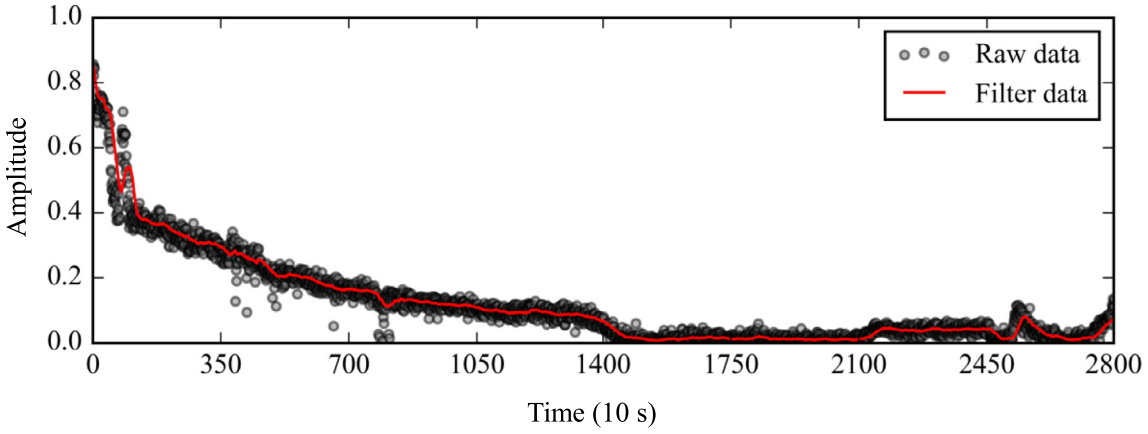


Fig. 8. Feature *F*2 of bearing1_1.

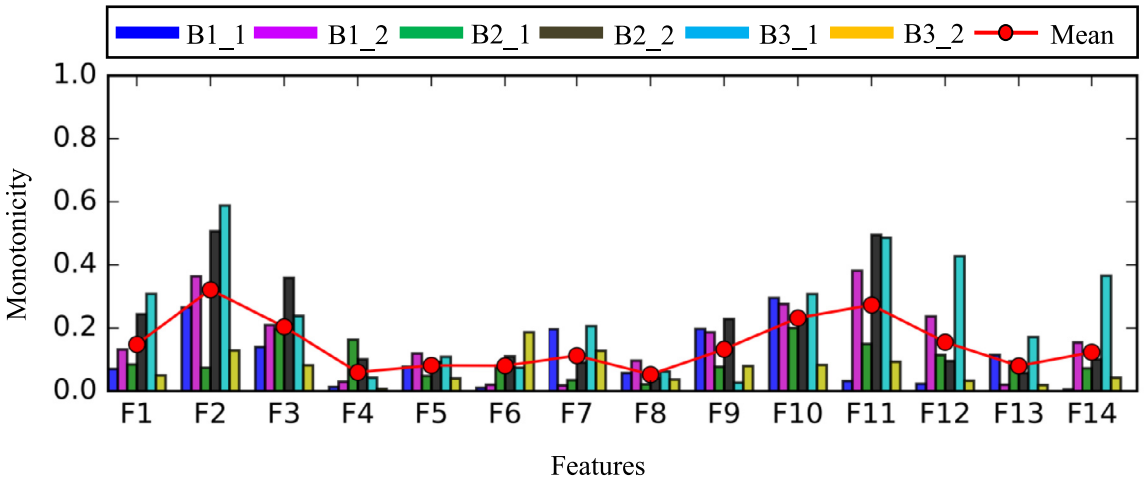


Fig. 9. Monotonicity of 14 features of six training bearings.

4.2. Health indicator construction

According to the construction procedure of the RNN-HI, 14 features are firstly extracted to construct a feature set. Then, the Mon and Corr metrics of the 14 features, as shown in Figs. 9 and 10, are compared on the six training bearings. For these two metrics, it is seen that the proposed frequency domain RS feature *F*2 holds the highest mean value. Furthermore, the feature *F*2 of bearing1_1 is plot in Fig. 8, where it is observed that the curve presents the distinct descent trend along with the degeneration of the bearing. In order to select the optimal feature subset, the criteria coefficients Cri of such extracted features are calculated and their normalized values are shown in Fig. 11, where *F*1, *F*2, *F*3, *F*9, *F*10, *F*11, *F*12 and *F*14 above a threshold of 0.5 are chosen as a feature subset. Specifically, three of these features are the proposed RS features, which demonstrate the effectiveness of the RS features for the assessment of bearing degradation process. To gain deeper insight into the selected features, we compare the performance of the selected features with that of some classical features. The results in Table 2 show that the classical features have lower Cri values and larger scales. On the contrary, the selected features have higher Cri values with the scales between 0 and 1. It means that the selected features can be employed without normalization.

Then, these selected features are input into the RNN to obtain the RNN-HI. The RNN-HI curves of six training bearings are shown in Fig. 12(a). In order to highlight the benefit of the RNN-HI, a

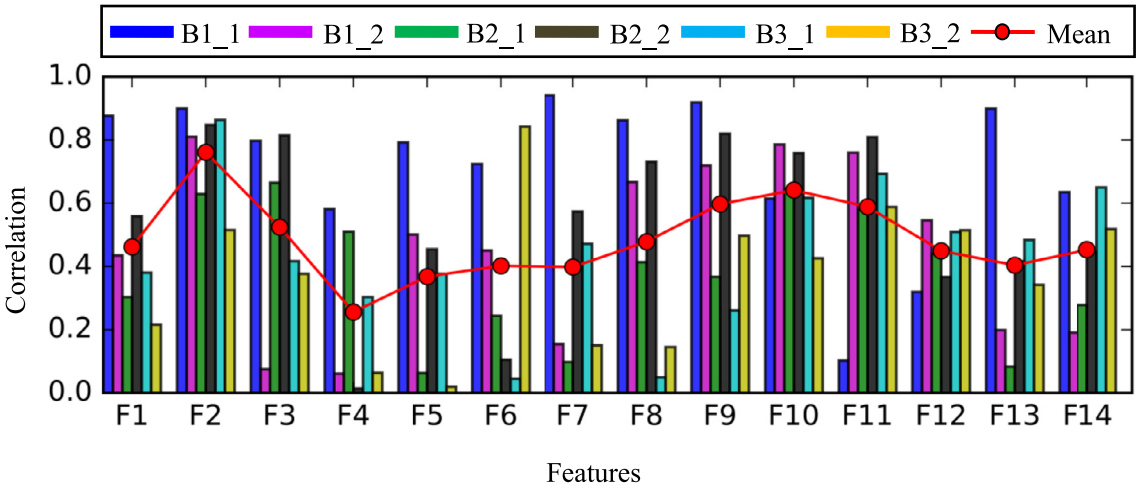


Fig. 10. Correlation of 14 features of six training bearings.

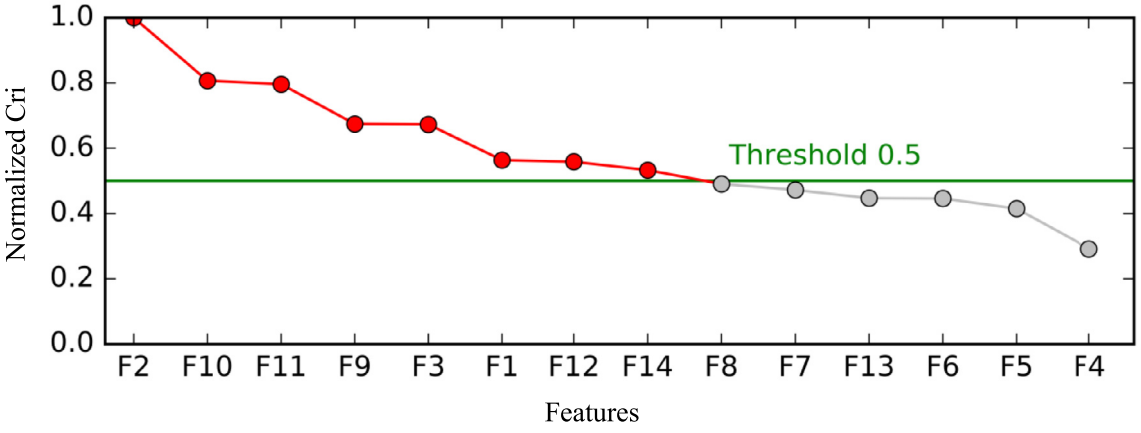


Fig. 11. Normalized Cri of 14 features of six training bearings.

Table 2
Metrics comparison of features.

Classical features				Selected features			
Feature	Cri	Min	Max	Feature	Cri	Min	Max
RMS	0.261	0.039	6.707	*F2*	0.5411	0.001	1.000
Wave factor	0.259	−1.192	4.041	*F10*	0.4367	0.001	0.829
Peak-to-peak	0.244	0.259	92.443	*F11*	0.4308	0.003	0.757
Margin factor	0.224	2.553	148.601	*F9*	0.3650	0.003	0.309
Variance	0.218	6.001	45.005	*F3*	0.3642	0.008	1.000
Kurtosis	0.173	2.545	415.831	*F1*	0.3050	0.004	1.000
Crest factor	0.157	2.359	17.722	*F12*	0.3023	0.007	0.296
Impulse factor	0.154	−38,855.215	70,059.176	*F14*	0.2883	0.003	0.202
Skewness	0.152	−14.409	1.0712				
Entropy	0.128	2.066	6.294				
Mean	0.021	−0.280	0.623				

self-organizing map based HI (SOM-HI) [36,37] is used for a comparison (see in Fig. 12(b)). Concretely, the Cri parameters of both HIs are higher than the selected features with 0.7471 for RNN-HI and 0.6877 for SOM-HI. Note that, although these two HI curves possess the similar Cri, their values at a failure condition appear to be different. In particular, whatever the operating condition of a bearing is, the RNN-HI is almost equal to one at the failure condition. However, the SOM-HI ranges from 0.54 to 1.75 for the six training bearings at the failure condition, which causes that the FT is difficult to determine.

To demonstrate the capability of the RNN-HI for RUL prediction, a double exponential model is built in this paper. The double exponential model has been demonstrated as an effective model for curve fitting and prediction [38], which is described as:

$$Y = ae^{bt} + ce^{dt} \tag{14}$$

where Y is the bearing's condition value; t is the time point; a, b, c and d are the model parameters.

Once the model is built, the RNN-HI is utilized to update its parameters and predict the RUL value based on the particle filtering algorithm [39]. Taking bearing2_6 as an example, a probability density function of 2000 particles represents the failure time distribution when the predicted RNN-HI reaches the failure threshold. As shown in Fig. 13, the medial value and 95% confidence

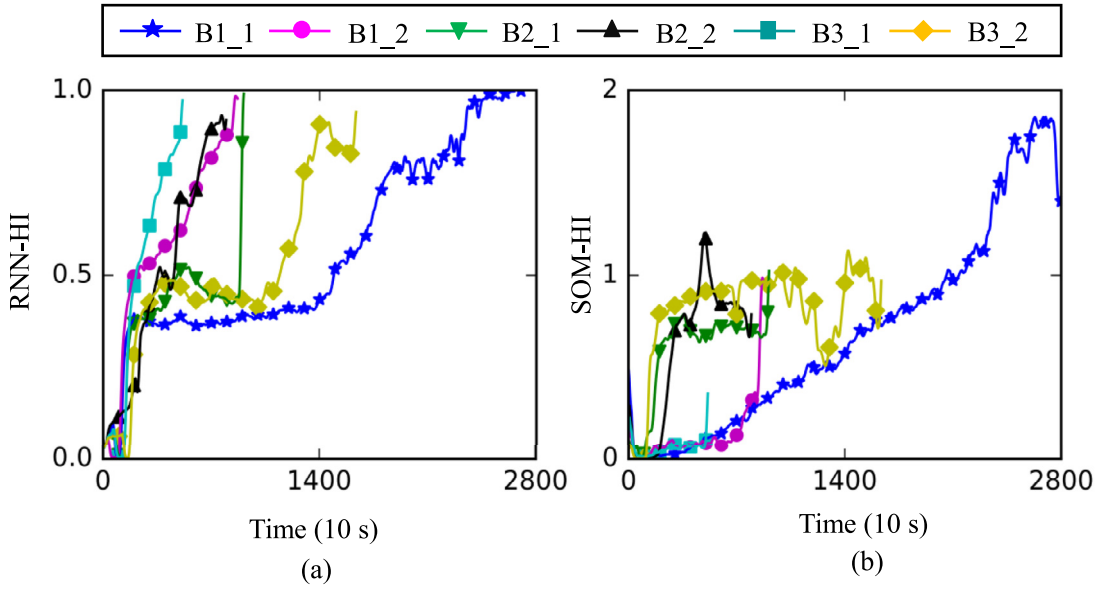


Fig. 12. HIs of the six training bearings: (a) the RNN-HI; (b) the SOM-HI.

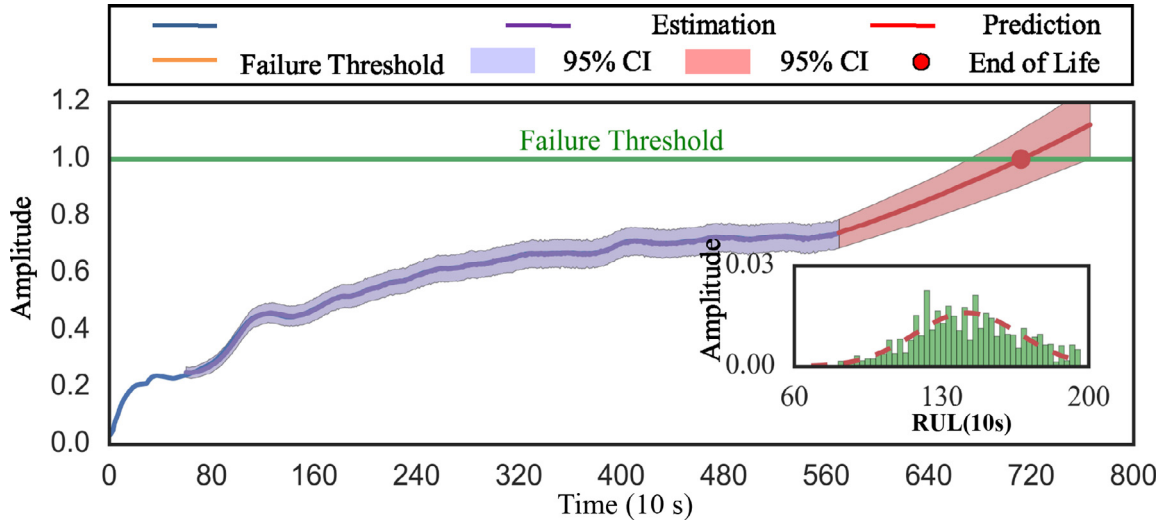


Fig. 13. RUL prediction result of bearing2_6.

Table 3
RUL prediction results.

Testing dataset	Current time (s)	Actually RUL (s)	Predict RUL (s)	RNN-HI (%)	SOM-HI (%)
Bearing1_3	18,010	5730	3250	43.28	−31.76
Bearing1_4	11,380	2900	1100	67.55	62.76
Bearing1_5	23,010	1610	1980	−22.98	−136.03
Bearing1_6	23,010	1460	1150	21.23	−32.88
Bearing1_7	15,010	7570	6220	17.83	−11.09
Bearing2_3	12,010	7530	4680	37.84	44.22
Bearing2_4	6110	1390	1660	−19.42	−55.40
Bearing2_5	20,010	3090	1410	54.37	68.61
Bearing2_6	5710	1290	1470	−13.95	−51.94
Bearing2_7	1710	580	900	−55.17	−68.97
Bearing3_3	3510	820	790	3.66	−21.96
Mean of *Er*				**32.48**	53.24

interval (CI) of the probability density function were 1470 s and [720 s, 1690 s], respectively.

The prediction results for the rest of the bearings are tabulated in Table 3. A percent error of prediction results is applied to evaluate the performance of the prediction method, which is defined as form:

$$Er_i = \frac{ActRUL_i - RUL_i}{ActRUL_i} \times 100\% \tag{15}$$

where $ActRUL_i$ and RUL_i are an actual RUL and an predicted one of the ith testing data, respectively.

Furthermore, the prediction results are compared with one similar research where HI construction is developed from the SOM [40]. Here, at three operating conditions, the FTs are chosen as 1.7, 1.3 and 1.1, respectively. The percent errors are shown in the column of 'SOM-HI'. From Table 3, the proposed method has the lower percent error than the SOM based method. This result strengthens the usability of the RNN-HI for RUL prediction.

5. Industrial data verification

In this section, the proposed RNN-HI is utilized to predict RUL of generator bearings from wind turbines. Concretely, as shown in Fig. 14, the driven-train of wind turbines was installed some accelerometers to monitor its health state. The sampling frequency

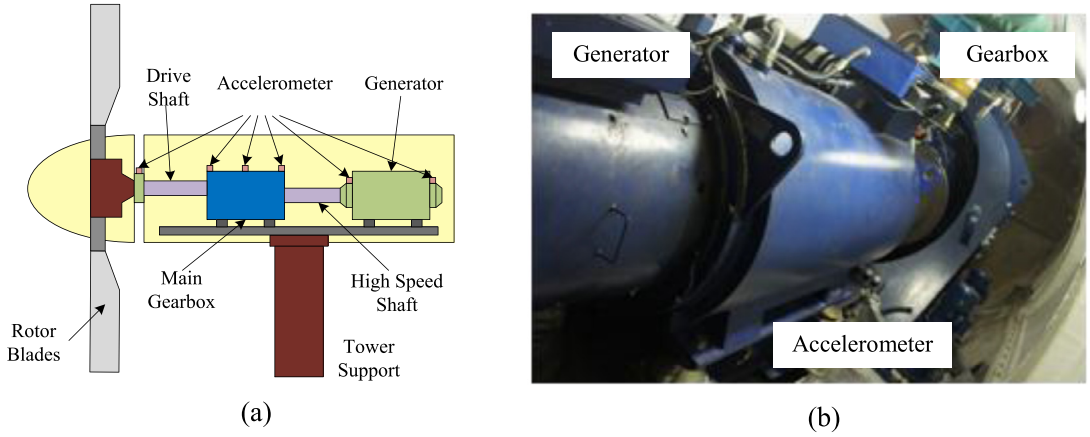


Fig. 14. A wind turbine: (a) a system structure and (b) a test system.

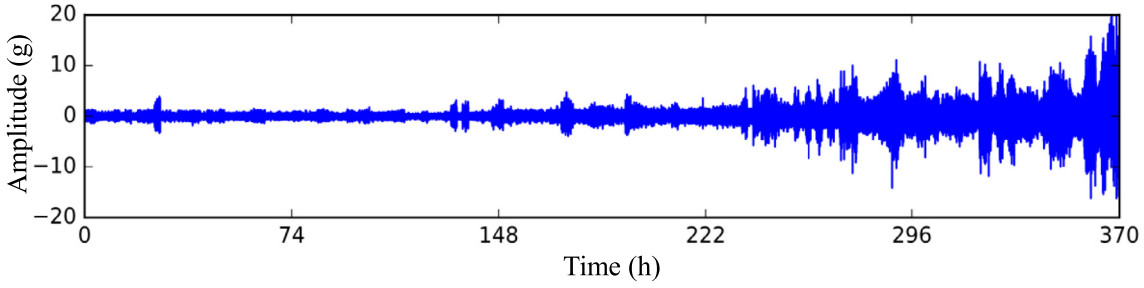


Fig. 15. The temporal vibration signal of FJ1.

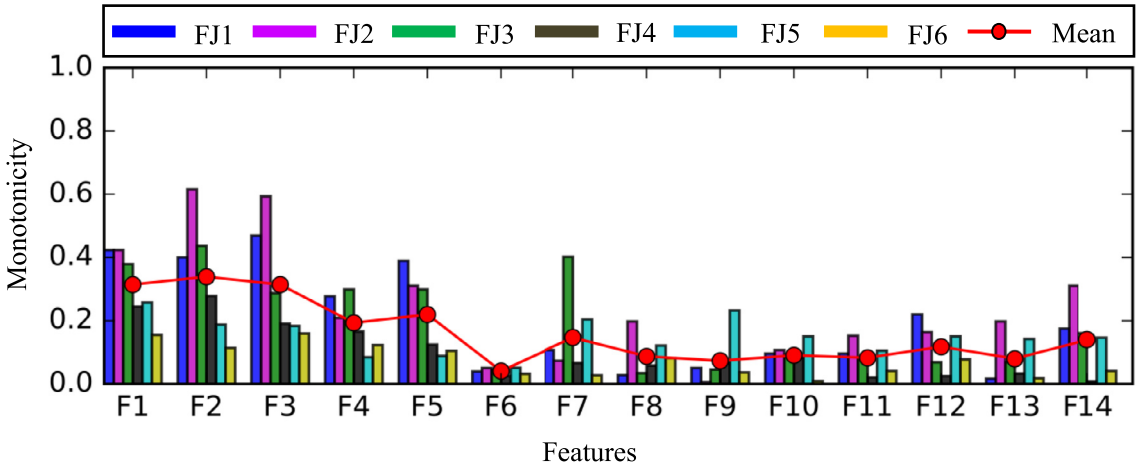


Fig. 16. Monotonicity of 14 features of six generator bearings.

of each accelerometer was set to 25.6 kHz, and the acquisition was conducted every 4 h The length of vibration signals is 4 s. In addition, a supervisory control and data acquisition (SCADA) system was also used to collect the operation information of wind turbines, such as rotating speed, active power, etc., which was helpful for the selection of the usable vibration signals. In the monitoring wind farms, from 25/09/2014 to 01/01/2016, there were six generator bearings which appeared to degenerate called FJ1, FJ2, FJ3, FJ4, FJ5 and FJ6, respectively. Fig. 15 shows the vibration signal of FJ1, and from the figure, the amplitude of vibration signal increases gradually with the time going on.

The Mon and Corr metrics of the extracted features $\{F_i\}_{i=1}^{14}$ are shown in Figs. 16 and 17, respectively. From the figures, the top four of these 14 features are all from the RS features, where feature $F2$ outperforms the others with the highest values of Mon and Corr. It means that the proposed features may have the properties to index the degradation process of generator bearings in this task. Then, the features whose normalized Cri cross the predefined threshold 0.5, $F1$, $F2$, $F3$, $F4$, $F5$, $F12$ and $F14$, are chosen as the feature subset to construct the RNN-HI. The average Cri of RNN-HIs is 0.7658 that is higher than all of the selected features. These results imply that the constructed RNN-HI maybe more suitable to describe the degradation processing of bearings than the selected single features.

In order to illustrate the performance of the RNN-HI for RUL prediction of generator bearings, a double exponential model is used in this case. For example, the RNN-HI of FJ1 is shown in Fig. 18. It is seen that the prediction curve roughly represents the degradation trend. Moreover, the RUL prediction based on SOM-HI is also conducted to compare with the proposed RNN-HI (see in

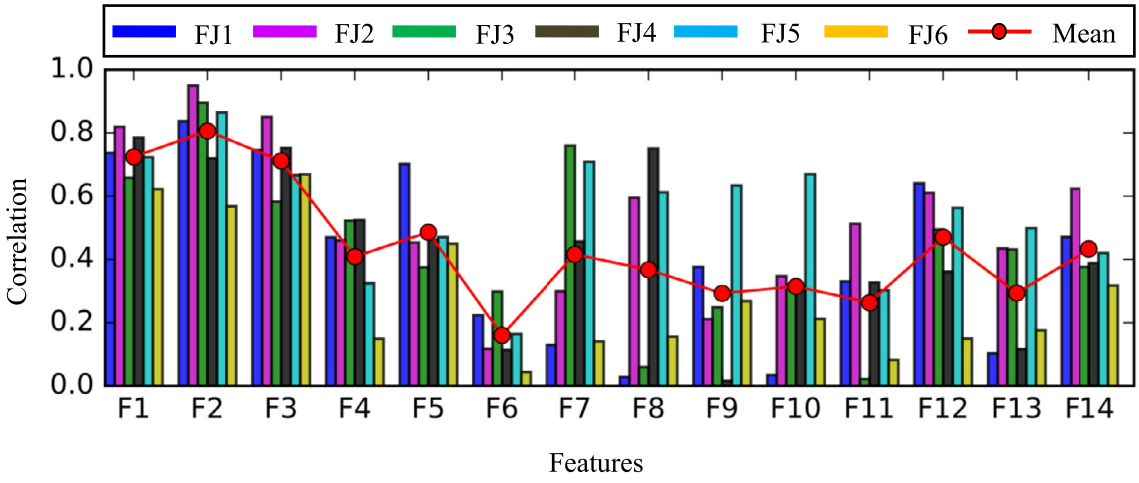


Fig. 17. Correlation of 14 features of six generator bearings.

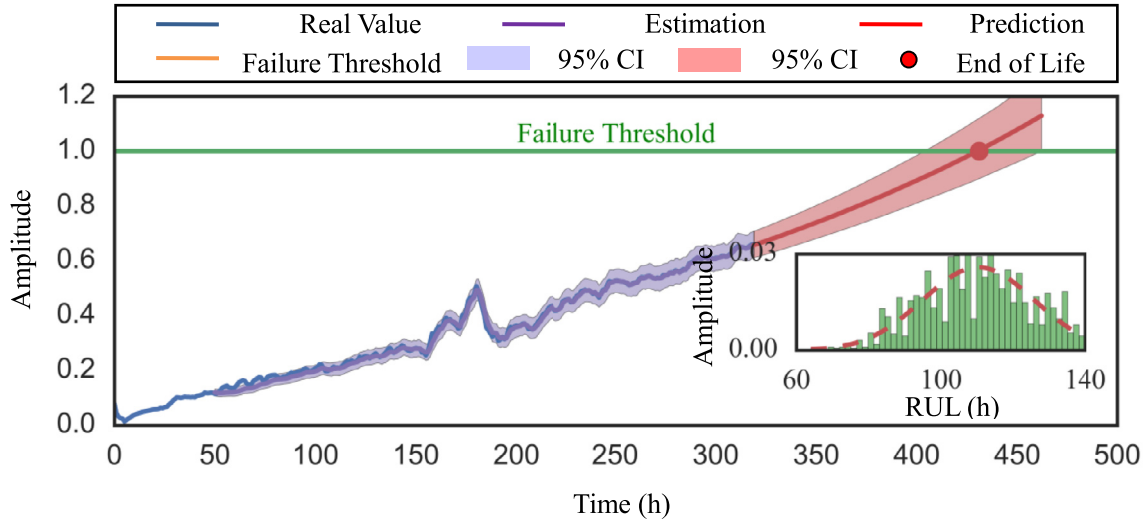


Fig. 18. RUL prediction result of FJ1.

Table 4
RUL prediction results.

Testing dataset	Current time (h)	Actually RUL (h)	Predict RUL (h)	RNN-HI (%)	SOM-HI (%)
FJ1	300	183	108	40.9	−210.35
FJ2	120	58	18	69	28.96
FJ3	135	40	38	5	−117.50
FJ4	130	48	46	4.1	−8.33
FJ5	320	166	124	0.25	−14.16
FJ6	330	109	131	−20.2	22.94.
Mean of *Er*				**23.24**	67.09

Table 4). According to the results, one can clearly notice that the RUL based on RNN-HI have lower average errors than the SOM-HI based method. These results further support the claims that the RNN-HI is useful for the RUL prediction of generator bearings.

6. Conclusions

RUL prediction accuracy highly depends on the performance of a HI. In this paper, the RNN-HI was proposed to enhance RUL prediction accuracy of bearings. During the construction procedure of the HI, the feature extraction method was proposed to map the classical time and frequency domain features with diversity ranges to some RS features ranging from 0 to 1. Subsequently, the most informative and sensitive features were selected based on the Cri metric. At last, these selected features were input into the RNN to construct the RNN-HI which fused the mutual information from the multiple features and properly correlated to the degradation processes of bearings. The vibration signals from the experimental platform and the wind turbine generators were collected to verify the proposed method. In the experiment verification, the SOM based method that used the same data sets was compared with the proposed method. The results showed that the proposed method performs better than the commonly used HI based on SOM. Moreover, the industrial data verification showed that the proposed method can predict the wind turbine generator bearing RUL effectively. In this paper, we focused on the construction of the HIs and then we used the double exponential model to validate the effectiveness of the proposed RNN-HI. So in future, RUL models, such as a conditional three-parameter capacity degradation model [41], a stochastic degradation model [42], etc., will be deeply studied so as to achieve higher prediction accuracies of RUL.

Acknowledgments

This research was supported by National Natural Science Foundation of China (51475355 and 61673311), Young Talent Support

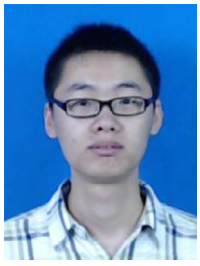
plan of Central Organization Department, and Visiting Scholar Foundation of the State Key Lab. of Traction Power in Southwest Jiaotong University (TPL1703).

References

[1] Y. Lei, J. Lin, M.J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: a review, Measurement 48 (2014) 292–305.
[2] A. Ghods, H.H. Lee, Probabilistic frequency-domain discrete wavelet transform for better detection of bearing faults in induction motors, Neurocomputing 188 (2016) 206–216.
[3] J. Liu, W. Wang, F. Golnaraghi, An enhanced diagnostic scheme for bearing condition monitoring, IEEE Trans. Instrum. Meas. 59 (2010) 309–321.
[4] Y. Wang, G. Xu, Q. Zhang, D. Liu, K. Jiang, Rotating speed isolation and its application to rolling element bearing fault diagnosis under large speed variation conditions, J. Sound Vib. 348 (2015) 381–396.
[5] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mech. Syst. Signal Process. 72 (2016) 303–315.
[6] J. Sikorska, M. Hodkiewicz, L. Ma, Prognostic modelling options for remaining useful life estimation by industry, Mech. Syst. Signal Process. 25 (2011) 1803–1836.
[7] D. Wang, C. Shen, An equivalent cyclic energy indicator for bearing performance degradation assessment, J. Vib. Control 22 (10) (2016) 2380–2388.
[8] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, J. Dybala, A model-based method for remaining useful life prediction of machinery, IEEE Trans. Reliab. 65 (2016) 1314–1326.
[9] N. Li, Y. Lei, J. Lin, S.X. Ding, An improved exponential model for predicting remaining useful life of rolling element bearings, IEEE Trans. Ind. Electron. 62 (2015) 7762–7773.
[10] D. Pan, J.-B. Liu, J. Cao, Remaining useful life estimation using an inverse Gaussian degradation model, Neurocomputing 185 (2016) 64–72.
[11] A. Heng, S. Zhang, A.C. Tan, J. Mathew, Rotating machinery prognostics: state of the art, challenges and opportunities, Mech. Syst. Signal Process. 23 (2009) 724–739.
[12] J. Liu, W. Wang, F. Ma, Y.B. Yang, C.S. Yang, A data-model-fusion prognostic framework for dynamic system state forecasting, Eng. Appl. Artif. Intell. 25 (2012) 814–823.
[13] K. Javed, A Robust & Reliable Data-Driven Prognostics Approach Based on Extreme Learning Machine and Fuzzy Clustering, Université de Franche-Comté, 2014.
[14] H. Qiu, J. Lee, J. Lin, G. Yu, Robust performance degradation assessment methods for enhanced rolling element bearing prognostics, Adv. Eng. Inform. 17 (2003) 127–140.
[15] D Wang, T Peter W, W Guo, et al., Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis, Meas. Sci. Technol. 22 (2) (2010) 025102.
[16] Y. Wang, Y. Peng, Y. Zi, X. Jin, K.-L. Tsui, A two-stage data-driven-based prognostic approach for bearing degradation problem, IEEE Trans. Ind. Inform. 12 (2016) 924–932.
[17] A. Widodo, B.-S. Yang, Machine health prognostics using survival probability and support vector machine, Expert Syst. Appl. 38 (2011) 8430–8437.
[18] F. Camci, K. Medjaher, N. Zerhouni, P. Nectoux, Feature evaluation for effective bearing prognostics, Qual. Reliab. Eng. Int. 29 (2013) 477–486.
[19] J. Yu, A hybrid feature selection scheme and self-organizing map model for machine health assessment, Appl. Soft Comput. 11 (2011) 4041–4054.
[20] H. Ocak, K.A. Loparo, F.M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics, J. Sound Vib. 302 (2007) 951–961.
[21] L. Ren, W. Lv, Remaining useful life estimation of rolling bearings based on sparse representation, in: Proceedings of the 7th International Conference on Mechanical and Aerospace Engineering (ICMAE), IEEE, 2016, pp. 209–213.
[22] E. Sutrisno, H. Oh, A.S.S. Vasan, M. Pecht, Estimation of remaining useful life of ball bearings using data driven methodologies„ in: Proceedings of the IEEE Conference on Prognostics and Health Management (PHM), IEEE, 2012, pp. 1–7.
[23] K. Medjaher, N. Zerhouni, J. Baklouti, Data-driven prognostics based on health indicator construction: application to PRONOSTIA's data, in: Proceeding of the European Control Conference (ECC), IEEE, 2013, pp. 1451–1456.
[24] J. Yu, Adaptive hidden Markov model-based online learning framework for bearing faulty detection and performance degradation monitoring, Mech. Syst. Signal Process. 83 (2017) 149–162.
[25] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Health assessment and life prediction of cutting tools based on support vector regression, J. Intell. Manuf. 26 (2015) 213–223.
[26] S.-j. Wu, N. Gebraeel, M.A. Lawley, Y. Yih, A neural network integrated decision support system for condition-based optimal predictive maintenance policy, IEEE Trans. Syst. Man. Cybern.-Part A: Syst. Hum. 37 (2007) 226–236.
[27] A. Cherif, H. Cardot, R. Boné, SOM time series clustering and prediction with recurrent neural networks, Neurocomputing 74 (2011) 1936–1944.
[28] P. Tse, D. Atherton, Prediction of machine deterioration using vibration based fault trends and recurrent neural networks, J. Vib. Acous. 121 (1999) 355–362.
[29] F.O. Heimes, Recurrent neural networks for remaining useful life estimation, in: Proceedings of the International Conference on Prognostics and Health Management (PHM), IEEE, 2008, pp. 1–6.
[30] A. Malhi, R. Yan, R.X. Gao, Prognosis of defect propagation based on recurrent neural networks, IEEE Trans. Instrum. Meas. 60 (2011) 703–711.
[31] Z.C. Lipton, J. Berkowitz, C. Elkan, A critical review of recurrent neural networks for sequence learning, arXiv preprint arXiv:1506.00019, 2015.
[32] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[33] K. Javed, R. Gouriveau, N. Zerhouni, P. Nectoux, A feature extraction procedure based on trigonometric functions and cumulative descriptors to enhance prognostics modeling, in: Proceedings of the IEEE Conference on Prognostics and Health Management (PHM), IEEE, 2013, pp. 1–7.
[34] J. Coble, J.W. Hines, Applying the general path model to estimation of remaining useful life, Int. J. Progn. Health Manag. 2 (2011) 71–82.
[35] 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 the IEEE Conference on Prognostics and Health Management (PHM), IEEE, 2012, pp. 1–8.
[36] R. Huang, L. Xi, X. Li, C.R. Liu, H. Qiu, J. Lee, Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, Mech. Syst. Signal Process. 21 (2007) 193–207.
[37] P. Sarlin, Self-organizing time map: an abstraction of temporal multivariate patterns, Neurocomputing 99 (2013) 496–508.
[38] X. Jin, Y. Sun, Z. Que, Y. Wang, T.W. Chow, Anomaly detection and fault prognosis for bearings, IEEE Trans. Instrum. Meas. 65 (2016) 2046–2054.
[39] Y.N. Qian, R.Q. Yan, Remaining useful life prediction of rolling bearings using an enhanced particle filter, IEEE Trans. Instrum. Meas. 64 (2015) 2696–2707.
[40] S. Hong, Z. Zhou, E. Zio, K. Hong, Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method, Digital Signal Process. 27 (2014) 159–166.
[41] D Wang, Q Miao, M Pecht, Prognostics of lithium-ion batteries based on relevance vectors and a conditional three-parameter capacity degradation model, J. Power Sources 239 (2013) 253–264.
[42] W Wang, X Liu, F Cai, et al., Stochastic dynamic modeling of lithium battery via expectation maximization algorithm, Neurocomputing 175 (2016) 421–426.

Liang Guo received the B.S. and Ph.D. degrees in mechanical engineering from Southwest Jiaotong University, Chengdu, China, in 2011 and 2016, respectively. He is currently working as a Postdoctoral researcher at the State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an, China. His current research interests include machinery condition monitoring, intelligent fault diagnostics and remaining useful life prediction.

Naipeng Li is currently working toward the Ph.D. degree in mechanical engineering at the State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, China. He received the B.S. degree in mechanical engineering from Shandong Agricultural University, China, in 2012. His research interests are machinery condition monitoring, intelligent fault diagnostics and remaining useful life prediction of rotating machinery.



Feng Jia is currently working toward the Ph.D. degree in mechanical engineering at the State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, China. He received the B.S. and M.S. degree in mechanical engineering from Taiyuan University of Technology, China, in 2011 and 2014, respectively. His research interests include machinery condition monitoring and fault diagnosis, intelligent fault diagnostics of rotating machinery.

Yaguo Lei received the B.S. and Ph.D. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2002 and 2007, respectively. He is currently a Full Professor of mechanical engineering at Xi'an Jiaotong University. Prior to joining Xi'an Jiaotong University in 2010, he was a Postdoctoral Research Fellow with the University of Alberta, Edmonton, AB, Canada. He was also an Alexander von Humboldt Fellow with the University of Duisburg-Essen, Duisburg, Germany. His research interests focus on machinery condition monitoring and fault diagnosis, mechanical signal processing, intelligent fault diagnostics, and remaining useful life prediction. Dr. Lei is a member of the editorial boards of more than ten journals, including Mechanical System and Signal Processing and Neural Computing & Applications. He is also a member of ASME and a member of IEEE. He has pioneered many signal processing techniques, intelligent diagnosis methods, and remaining useful life prediction models for machinery.

Jing Lin received his B.S., M.S. and Ph.D. degrees from Xi'an Jiaotong University, China, in 1993, 1996 and 1999, respectively, all in mechanical engineering. He is currently a Professor with the State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University. From July 2001 to August 2003, he was a Postdoctoral Fellow with the University of Alberta, Edmonton, AB, Canada, and a Research Associate with the University of Wisconsin–Milwaukee, Milwaukee, WI, USA. From September 2003 to December 2008, he was a Research Scientist with the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, under the sponsorship of the Hundred Talents Program. His current research directions are in mechanical system reliability, fault diagnosis, and wavelet analysis. Dr. Lin was a recipient of the National Science Fund for Distinguished Young Scholars in 2011.