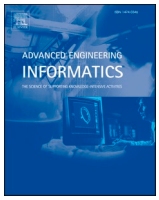
Advanced Engineering Informatics 53 (2022) 101682

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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



Full length article

Remaining useful life prediction of bearings by a new reinforced memory GRU network

Jianghong Zhou, Yi Qin *, Dingliang Chen, Fuqiang Liu, Quan Qian

State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, People's Republic of China
College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, People's Republic of China

ARTICLE INFO

Keywords:
Attention mechanism
Gated recurrent unit
Health indicator
Reinforced memory
RUL prediction

ABSTRACT

The remaining useful life (RUL) prediction of bearings has great significance in the predictive maintenance of mechanical equipment. Owing to the difficulty of collecting abundant lifecycle datasets with correct labels, it is quite necessary to explore a prediction method with high precision and robustness in the case of small samples. It follows that a novel RUL prediction approach is put forward to overcome this problem. First, for reducing the man-made interference and the demand for expert knowledge, an unsupervised health indicator (HI) is constructed by Gaussian mixture model (GMM) and Kullback-Leibler divergence (KLD), which is named as KLD-based HI. Then because of the rapid forgetting of historical trend information in the current RNN-based prediction models, a novel reinforced memory gated recurrent unit (RMGRU) network is proposed by reusing the state information at the previous moment. According to the constructed KLD-based HI vector, the unknown HIs are successively predicted by RMGRU until the predicted HI value exceeds the failure threshold, and then RUL is calculated. The contrast experiment on IEEE 2012PHM bearing datasets shows the superiority of the bearing RUL prediction approach based on RMGRU over the classical time series forecasting methods. It can be concluded that this method has great application potential in bearing RUL prediction.

1. Introduction

With the fast development of Industry 4.0, prognostics health management (PHM) technology with fault prognosis as the core has aroused extensive attention [1]. Rolling bearings are widely utilized in electric motors, turbofan engines, machine tools and other complex mechanical equipment [2]. Unfortunately, rolling bearings are prone to failure in complex and harsh working conditions. Therefore, in order to make the optimal maintenance decision and enhance the reliability of mechanical equipment operation, it is imperative to predict the RULs of bearings. At present, the RUL prediction approaches can be roughly categorized as model-based, data-driven and hybrid methods [3,4].

The model-based methodology needs expert knowledge to build an exact mathematical model for exploring the degradation trajectory of machinery [5,6]. Under the condition of accurate modeling, these methods have high RUL prediction precision. However, as the complexity of mechanical systems increases, it is more difficult to construct an accurate mathematical model, and the disturbances and various uncertainties in the real systems will also affect the accuracy of the modeling [7,8]. As a result, the application potential of model-based approaches is limited. Data-driven methodologies need to collect the monitoring data of equipment, and then build an end-to-end model to learn the hidden laws. It can mine the nonlinear relationship between the input data and the target without specific physical knowledge and expert knowledge [9]. Therefore, this kind of methodology can overcome the difficulties of model-based methodology in the field of machinery RUL prediction. The hybrid methodologies are formed by combining the above two approaches. The failure mechanism of the mechanical equipment is utilized to establish the mathematical model, after that the data-driven technique is employed to update the model parameters [10,11]. Although hybrid approaches have a more precise result of prediction, it still requires an accurate mathematical model. Comparing three types of RUL prediction methods, the data-driven methodology is most widely used, and it is least affected by the expertise.

The data-driven methodology can be roughly divided into pattern recognition methods and time series prediction methods [12,13]. The pattern recognition approaches were used to predict RULs by directly building the mapping functions between the input (multi-sensory data or multi-features) and the output (RUL) [14,15]. The construction of

* Corresponding author at: State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, People's Republic of China.
E-mail address: qy_808@cqu.edu.cn (Y. Qin).

https://doi.org/10.1016/j.aei.2022.101682
Received 1 March 2022; Received in revised form 13 May 2022; Accepted 25 June 2022
Available online 1 July 2022

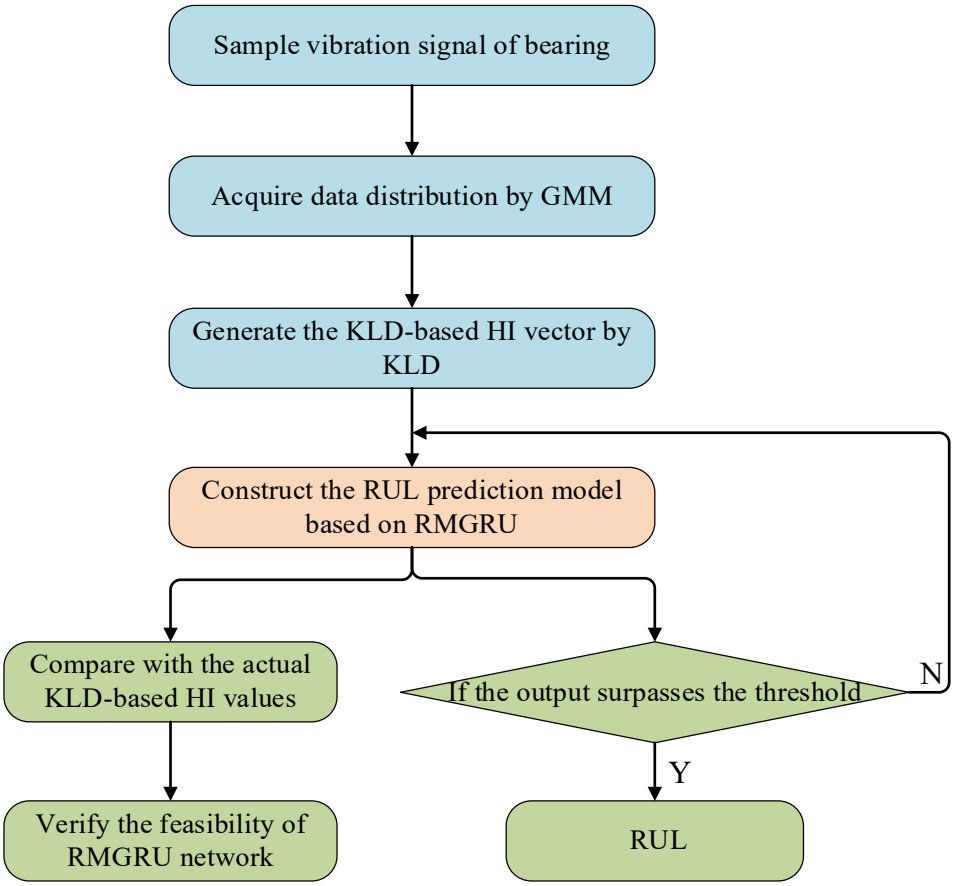


Fig. 1. The proposed bearing RUL prediction flowchart.

mapping functions needs massive amounts of sensor data with labels. And in order to obtain a high precision prediction result, this kind of method requires similar degradation trajectories between different life-cycle datasets. However, in practical engineering, it is difficult to collect the massive and complete life-cycle datasets with similar degradation trends. Whereupon, the time series prediction approaches were proposed to overcome the issues of limited and unlabeled samples [12,16]. For this type of method, it is necessary to construct a HI that can reflect the degradation trajectory. With the collected samples, the corresponding HI points are constructed, then they can be modeled by the deep neural network or some machine learning approaches [18,19] such as polynomial regression, support vector regression, naive forecast, and autoregressive integrated moving average model, et al. Due to the strong nonlinear mapping capabilities of deep neural networks, they can be better applied to model the complex nonlinear HI data. Finally, the bearing RUL is estimated by the step-by-step prediction of HI and the failure threshold. For the time series prediction methods, the key is to establish a powerful neural network that can well learn the trend characteristics of HI.

As a kind of neural network specialized in processing sequence data, recurrent neural network (RNN) can use the previous output data and the current input data. For this reason, using RNN is more conducive to learning degradation trends from HI. Currently, Long short-term memory (LSTM) [20] and gated recurrent unit (GRU) [21] neural networks are two popular RNN structures. Compared to LSTM, GRU has fewer parameters which can reduce the consumption of computing resource. Owing to the avoidance of long-term dependence, LSTM and GRU have been utilized in various machines, such as aero-engine [22], turbofan engine [23], gear [24], bearing [25], and so on.

Attention mechanism can assign different weights to information according to its importance, which has the ability to reduce the information redundancy and rationally utilize the computing resource. The RNNs based on attention mechanism have received a lot of concerns in natural language processing (NLP) [26] and image identification [27]. With regard to RUL prediction, Chen et al. combined attention mechanism and encoder-decoder framework based on GRU to predict the RULs of bearings [28]. Chen et al. proposed a novel aero-engine RUL prediction framework via attention-based deep learning, which used the attention mechanism to learn the importance of LSTM's output at each moment [29]. Li et al. designed an attention module, which was embedded into the convolution LSTM network for RUL prediction [30].

The current RNN-based RUL prediction models are mainly developed by changing the cell structure [22] and the transition [31] of RNN, which do not make full of the historical state information. As a result, most RNN-based models will quickly forget the historical trend information when they are used to predict the RULs of bearings. Considering the rapid forgetting of historical trend information in the RNN-based models, a reinforced memory GRU (RMGRU) is developed by fully considering the state information at the previous two moments, then the attention mechanism is utilized to capture the important trend characteristics from the state information at the previous two moments and allocate the computing resource to them. Owing to this improvement,

RMGRU can enhance the learning ability of the trend characteristics of HI and increase the accuracy of RUL prediction. In addition, using Gaussian mixture model (GMM) and Kullback-Leibler divergence (KLD), an unsupervised HI named KLD-based HI is constructed that can well represent the degradation trend of bearings. Combing KLD-based HI and RMGRU, a novel time series prediction method is proposed, which uses the KLD-based HI points of some known samples in one dataset to predict the unknown KLD-based HI points by RMGRU. When the predicted HI value is larger than the predetermined failure threshold, the bearing RUL is calculated through the predicted points and the recording interval. The comparative experiments show that it has more exact and robust prediction results than the current typical time series prediction methods. In this study, the main contributions are given below:

(1) To better mine the degradation trend, a novel RMGRU network with attention mechanism is proposed to slow down the forgetting speed in RNN and its variants via reusing the state information at the previous moment.

(2) For reducing the man-made interference and the requirement of expert knowledge, an unsupervised KLD-based HI is constructed from the raw vibration samples by using GMM and KLD.

(3) A new time series RUL prediction approach is proposed via the KLD-based HI and RMGRU. Experiments on the IEEE 2012 PHM bearing datasets indicate that it has better predictive performance than the existing typical RUL forecasting methods.

The rest of this research is arranged as follows: In Section 2, we present a novel RUL prediction framework of bearing, including HI construction, RMGRU network and bearing RUL prediction. In Section 3, the contrast experiment is implemented to prove the superiority of the proposed methodology. Finally, the conclusions are drawn in Section 4.

2. Proposed framework for bearing RUL prediction

This section presents the proposed bearing RUL prediction framework based on the times series forecasting principle. The whole predictive procedure is illustrated in Fig. 1, which consists of HI construction and RUL prediction. First, an unsupervised HI with a distinct degradation trend is extracted by Gaussian mixture model (GMM) and Kullback-Leibler divergence (KLD) from the raw vibration signals. Then RMGRU is developed and applied to predict the KLD-based HI points step by step. Once the predicted HI value exceeds the predetermined failure threshold, the RUL of bearing is computed immediately.

2.1. HI construction based on GMM and KLD

For the data-driven method, a proper HI construction affects the performance of RUL prediction to a great degree [32]. The existing HI construction techniques are mainly classified into three categories, including signal processing [33], multi-feature fusion [34] and supervised learning [35]. The HI extracted by signal processing has a specific physical meaning, but a single feature cannot comprehensively reflect the degradation stage of bearing. The multi-feature fusion approaches rely on the extracted features and expert knowledge in significant measure. Although the supervised learning methods can mine the useful degraded information from the raw data to construct the HI, they need some suitable labels to constrain the degradation trajectory. Therefore, it is worth constructing an unsupervised HI by using the raw vibration samples in the actual engineering. With the performance degradation of bearing, the distributions of monitoring samples will change. Therefore, the distribution discrepancy can be used to construct the HI without supervision.

2.1.1. Gaussian mixture model

The aim of GMM is to describe a data sample distribution by a weight sum of multiple Gaussian components. Since the bearing vibration sample generally has Gaussian characteristics [36], GMM can be employed to estimate its distribution. Given a k-dimensional data sample$\boldsymbol{x} = (x^1, x^2, \cdots, x^k)$, it consists of M Gaussian components. Then according to Ref [37], GMM is defined as:

$$P(x^i|\boldsymbol{\Phi}) = \sum_{m=1}^{M} \pi_m p_m(x^i|\mu_m, \sigma_m) \tag{1}$$

where $P(x^i|\boldsymbol{\Phi})$ is the Gaussian mixture density, all the parameters of GMM are represented as$\boldsymbol{\Phi} = (\pi_1, \cdots, \pi_M; \mu_1, \cdots, \mu_M; \sigma_1, \cdots, \sigma_M)$, and π_m denotes the weight of p_m that satisfies$\sum_{m=1}^{M}\pi_m = 1$. For the bearing signal, each Gaussian component $p_m(x^i|\mu_m, \sigma_m)$ is an univariate Gaussian function given by:

$$p_k(x^i|\mu_m, \sigma_m) = \frac{1}{\sqrt{2\pi}\sigma_m}\exp\left\{-\frac{(x^i-\mu_m)^2}{2\sigma_m^2}\right\} \tag{2}$$

where μ_m and σ_m denote the mean and standard deviation respectively.

The maximum likelihood estimate technique is utilized to find the parameter$\boldsymbol{\Phi}$, and the log-likelihood function is calculated by:

$$\mathrm{G}(\boldsymbol{\Phi}) = \log P(\boldsymbol{x}|\boldsymbol{\Phi}) = \sum_{i=1}^{k}\log\left(\sum_{m=1}^{M}\pi_m p_m(x^i|\mu_m, \sigma_m)\right) \tag{3}$$

Since directly optimizing the Eq. (3) is intractable, the category control variable $z = (z_1, z_2, \cdots, z_M)$ is introduced into Eq. (3). It then follows that Eq. (3) can be rewritten as.

$$\begin{aligned}\mathrm{G}(\boldsymbol{\Phi}) &= \log P(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{\Phi}) = \log\prod_{i=1}^{k} P(x^i, z_j|\boldsymbol{\Phi}) \\ &= \sum_{i=1}^{k}\log P(x^i|z_j)P(z_j) = \sum_{i=1}^{k}\log\pi_{z_j}p_{z_j}(x^i|\mu_{z_j}, \sigma_{z_j})\end{aligned} \tag{4}$$

The above equation can be optimized by the famous Expectation-Maximization algorithm [38].

2.1.2. Kullback-Leibler divergence

KLD [39], namely relative entropy, was first introduced by Solomon Kullback and Richard Leibler in 1951. At present, it has been widely used in fault diagnosis, degradation assessment and other fields. It is usually used to measure the degree of discrepancy between two probability distributions. If the two distributions are the same, the value of KLD is 0. And as the difference between two probability distributions increases, so does the value of KLD. Consequently, KLD can be utilized for evaluating the discrepancy between two distributions. Assume that two arbitrary probability density functions of random variable Y are respectively denoted as $A(y)$ and $B(y)$. Then the KLD of $B(y)$ from $A(y)$ is defined as:

$$D_{KLD}(B(y)||A(y)) = \int B(y)log\frac{B(y)}{A(y)}dx \tag{5}$$

As previously mentioned, the distribution of bearing vibration

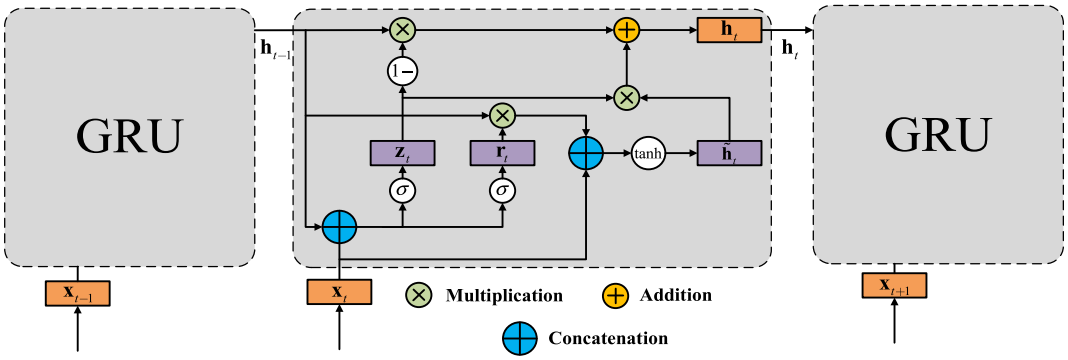


Fig. 2. The structure of GRU network.

sample is approximately subject to Gaussian distribution. Assume that $A(y)$ and $B(y)$ are respectively denoted as $N_1(\mu_1,\sigma_1)$ and$N_2(\mu_2,\sigma_2)$, the KLD of N_1 from N_2 can be calculated by:

$$\begin{aligned} &D_{KLD}(N_1(\mu_1,\sigma_1)||N_2(\mu_2,\sigma_2)) \\ &= \int\left(\frac{1}{\sqrt{2\pi}\,\sigma_1}\exp(-(x-\mu_1)^2/2\sigma_1^2)\right) log\left(\frac{\exp(-(x-\mu_1)^2/2\sigma_1^2)/\sqrt{2\pi}\,\sigma_1}{\exp(-(x-\mu_2)^2/2\sigma_2^2)/\sqrt{2\pi}\,\sigma_2}\right)dx \\ &= \int\left(\frac{1}{\sqrt{2\pi}\,\sigma_1}\exp(-(x-\mu_1)^2/2\sigma_1^2)\right)\left(\log\frac{\sigma_2}{\sigma_1}-\frac{(x-\mu_1)^2}{2\sigma_1^2}+\frac{(x-\mu_2)^2}{2\sigma_2^2}\right)dx \\ &= \frac{1}{2}(\log\frac{\sigma_2^2}{\sigma_1^2}-1+\frac{\sigma_1^2+(\mu_2-\mu_1)^2}{\sigma_2^2}) \end{aligned} \tag{6}$$

where log denotes the natural logarithm.

2.1.3. The procedure of KLD-based HI construction

The probability distribution p_0 of all healthy samples is taken as the baseline. When the bearing is healthy, the probability distribution of the collected sample is almost similar to p_0. Once the bearing degradation occurs, the probability distribution of the collected sample is different from p_0, and the difference will generally increase with the degree of degeneration. Therefore, this difference between p_0 and the probability distribution of the monitoring signal can be used to represent the health status of bearing, and it can be calculated by KLD. Based on the KLD of two univariate Gaussian distributions, we propose the KLD-based HI, and its construction procedure is summarized as follows:

Step1: The run-to-fail vibration dataset of bearings is collected. $\boldsymbol{X} = (\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_K)$ represents the acquired dataset, where$\boldsymbol{x}_i = (x^1,x^2,\cdots,x^k)$, K and k represent the number of samples and the length of a sample, respectively.

Step2: The first H samples $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_H$ of dataset $\mathbf{X}$ are treated as the baseline health samples and form a dataset, which is used to estimate the probability distribution p_0 by GMM. H is empirically set, and the number of Gaussian components in GMM is set as 4 [40].

Step3: The probability distributions $\boldsymbol{p} = [p_{H+1},\cdots,p_K]$ of the remaining samples $\boldsymbol{x}_{H+1},\cdots,\boldsymbol{x}_K$ are separately estimated by GMM with 4 Gaussian components.

Step4: The KLD-based HI is constructed. According to Eq. (6), the HI vector is generated by calculating the KLD between p_0 and$\boldsymbol{p} = [p_{H+1},\cdots,p_K]$.

2.2. RMGRU for the RUL prediction method

2.2.1. GRU neural network

Compared with LSTM, GRU has two major improvements: (1) it combines the unit state and output into the hidden state and only relies on the hidden state to transmit information; (2) it integrates the forget and input gates in LSTM into an update gate. Owing to the above two improvements, GRU has fewer parameters, which can effectively reduce the operation time. Meanwhile, the GRU network inherits the ability of LSTM to conquer the problems of vanishing gradient and gradient explosion. And its gate structure can effectively filter out the useless information. The structure of GRU network is depicted in Fig. 2, where σ denotes the sigmoid activation function and tanh denotes the hyperbolic tangent activation function.

At the time t, the update rule of hidden state $\boldsymbol{h}_t$ can be written as.

$$\boldsymbol{z}_t = \sigma(\boldsymbol{A}_z[\boldsymbol{x}_t,\boldsymbol{h}_{t-1}] + \boldsymbol{d}_z) \tag{7}$$

$$\boldsymbol{r}_t = \sigma(\boldsymbol{A}_r[\boldsymbol{x}_t,\boldsymbol{h}_{t-1}] + \boldsymbol{d}_r) \tag{8}$$

$$\widetilde{\boldsymbol{h}}_t = \tanh(\boldsymbol{A}_h[\boldsymbol{x}_t,\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}] + \boldsymbol{d}_h) \tag{9}$$

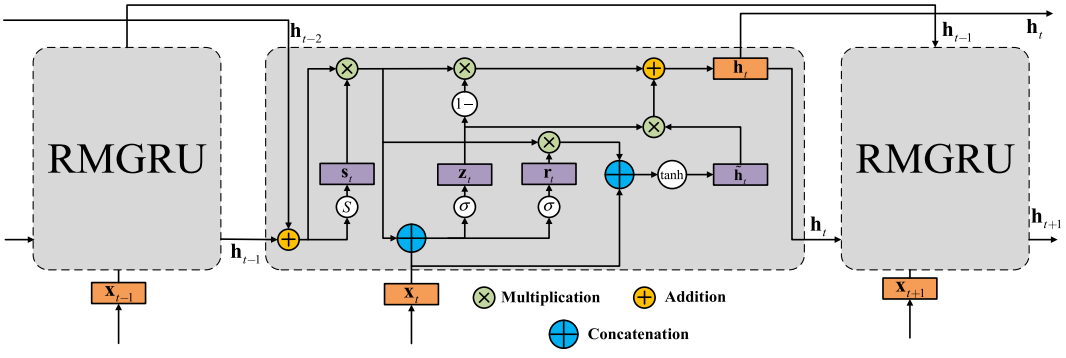


Fig. 3. The structure of RMGRU network.

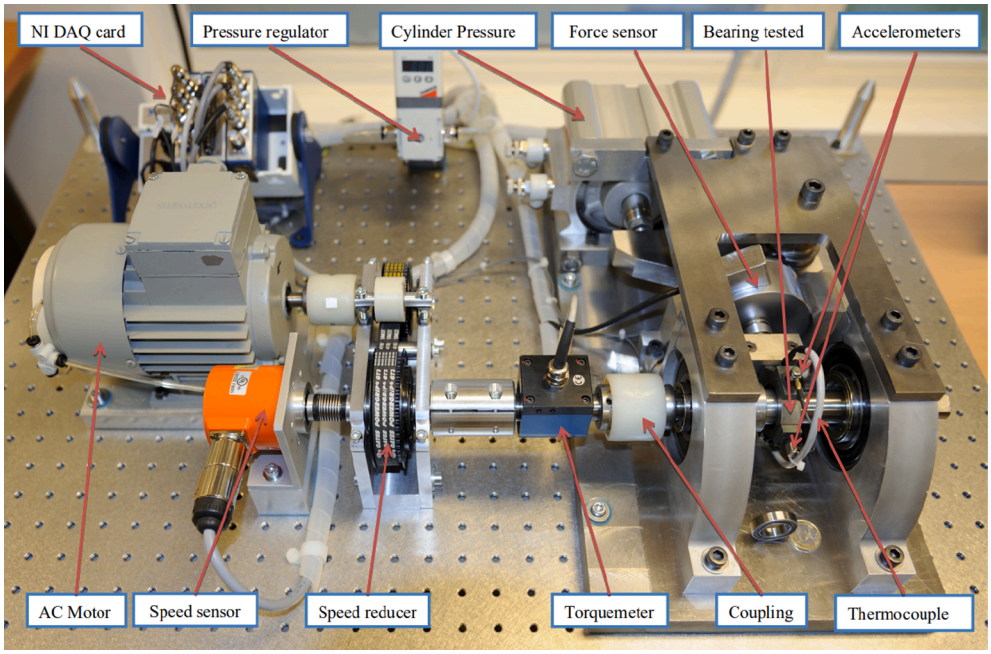


Fig. 4. PRONOSTIA platform.

$$\boldsymbol{h}_t = (1 - \boldsymbol{z}_t) \odot \boldsymbol{h}_{t-1} + \boldsymbol{z}_t \odot \widetilde{\boldsymbol{h}}_t \tag{10}$$

where $\boldsymbol{x}$ represents the input information; $\boldsymbol{h}$ is the hidden state and $\widetilde{\boldsymbol{h}}$ is the temporary hidden state; **A** is weight matrix and **d** is the bias matrix; **r** and **z** denote the output of reset gate and update gate respectively; and $\odot$ denotes dot product.

2.2.2. Attention mechanism

In order to get the needed information, the human brain tends to focus on a few important areas, and then establish the description of the environment [41]. Thereupon, the attention mechanism was put forward for mining the important characteristics. Assume that the input data is$\boldsymbol{x} = [x_1, x_2, \cdots, x_k]$, where k is the data length. The attention coefficient of the i^{th} input data x_i is computed by:

$$\alpha_i = \frac{\exp(s(x_i, q))}{\sum_{j=1}^{k} \exp(s(x_j, q))} \tag{11}$$

where q is the query vector that is related to the task. $s(x_i, q)$ is a score function, and its formula is written as:

$$s(x_i, q) = \boldsymbol{V}^T \tanh(\boldsymbol{W} x_i + \boldsymbol{U} q) \tag{12}$$

where **U**, **V**, **W** are the network parameters obtained by training the neural network.

2.2.3. Reinforced memory GRU network

When predicting the RUL of bearings, RNN will use the historical trend information in addition to the current state information. This is similar to the mechanism of human memory. As a result, the historical memory of RNN is very important for processing the long sequential series. The existing RNN-based models, such as GDAU [10] and DCLSTM [31], only improve the cell structure of RNN, but underutilize the historical trend information. It is revealed by H. Ebbinghaus that the oblivion begins when the information enters the brain, and the rate of oblivion decreases with time, especially in a short time after remembering, the forgetting rate is the largest [42]. Assume that the forgetting rule of RNN is similar to this forgetting rule revealed by H. Ebbinghaus. Inspired by the forgetting law, a reinforced memory GRU network is built to strengthen the learning capacity of tendency characteristics. Fig. 3 demonstrates the structure of RMGRU network. In this figure, **x** and **h** respectively represent the input data and hidden state, S represents Eq. (11) and $\mathbf{s}_t$ is the output of Eq. (11).

To enhance the capacity of long-term memory and obtain more robust predicting results, the hidden state containing the learned knowledge needs to be further memorized. On the basis of the forgetting law, the timely consolidation of learned knowledge can slow down the memory loss, hence the long-term memory ability of RNN can be enhanced by combining the state information at the previous two moments. As the direct usage of the historical status information at the previous two times will lead to information redundancy and increase the computing burden of networks, the attention mechanism is used for paying attention to the trend characteristics of HI points and allocate the computing resource to them. As is shown in Fig. 3, $\mathbf{h}_{t\text{-}2}$ and $\mathbf{h}_{t\text{-}1}$ are added, then the attention mechanism is used to capture the important trend characteristics from the state information at the previous two moments. From Fig. 3, the output hidden state at the time t is updated by:

$$\begin{aligned}
\overline{\boldsymbol{h}}_{t-1} &= \boldsymbol{h}_{t-1} + \boldsymbol{h}_{t-2} \\
s(\boldsymbol{h}_{t-1}, \boldsymbol{h}_{t-2}) &= \boldsymbol{V}^T \tanh(\boldsymbol{W}_s \boldsymbol{h}_{t-1} + \boldsymbol{U}_s \boldsymbol{h}_{t-2}) \\
\boldsymbol{\alpha}_{t-1} &= \frac{\exp(s(\boldsymbol{h}_{t-1}, \boldsymbol{h}_{t-2}))}{\sum_{j=1}^{n} \exp(s(\boldsymbol{h}_{t-1}, \boldsymbol{h}_{t-2}))} \\
\boldsymbol{\Omega}_{t-1} &= \boldsymbol{\alpha}_{t-1} \odot \overline{\boldsymbol{h}}_{t-1} \\
\boldsymbol{z}_t &= \sigma(\boldsymbol{A}_z [\boldsymbol{x}_t, \boldsymbol{A}_{t-1}] + \boldsymbol{d}_z) \\
\boldsymbol{r}_t &= \sigma(\boldsymbol{A}_r [\boldsymbol{x}_t, \boldsymbol{A}_{t-1}] + \boldsymbol{d}_r) \\
\widetilde{\boldsymbol{h}}_t &= \tanh(\boldsymbol{A}_h [\boldsymbol{x}_t, \boldsymbol{r}_t \odot \boldsymbol{\Omega}_{t-1}] + \boldsymbol{d}_h) \\
\boldsymbol{h}_t &= (1 - \boldsymbol{z}_t) \odot \boldsymbol{\Omega}_{t-1} + \boldsymbol{z}_t \odot \widetilde{\boldsymbol{h}}_t
\end{aligned} \tag{13}$$

It can be noted from Eq. (13) that the working of RMGRU is mainly divided into two parts. The first part is that the attention mechanism is used to guide the state information at the previous two moments for enhancing the memory ability of network. It first employs Eqs. (11) and (12) to focus on the hidden states $\mathbf{h}_{t\text{-}2}$ and $\mathbf{h}_{t\text{-}1}$, and calculates the attention coefficient, then the vital information $\boldsymbol{\Omega}_{t-1}$ can be captured from $\overline{\boldsymbol{h}}_{t-1}$ which is fused by $\mathbf{h}_{t\text{-}2}$ and $\mathbf{h}_{t\text{-}1}$ through the attention coefficient. The second part is to mine the degradation information from the

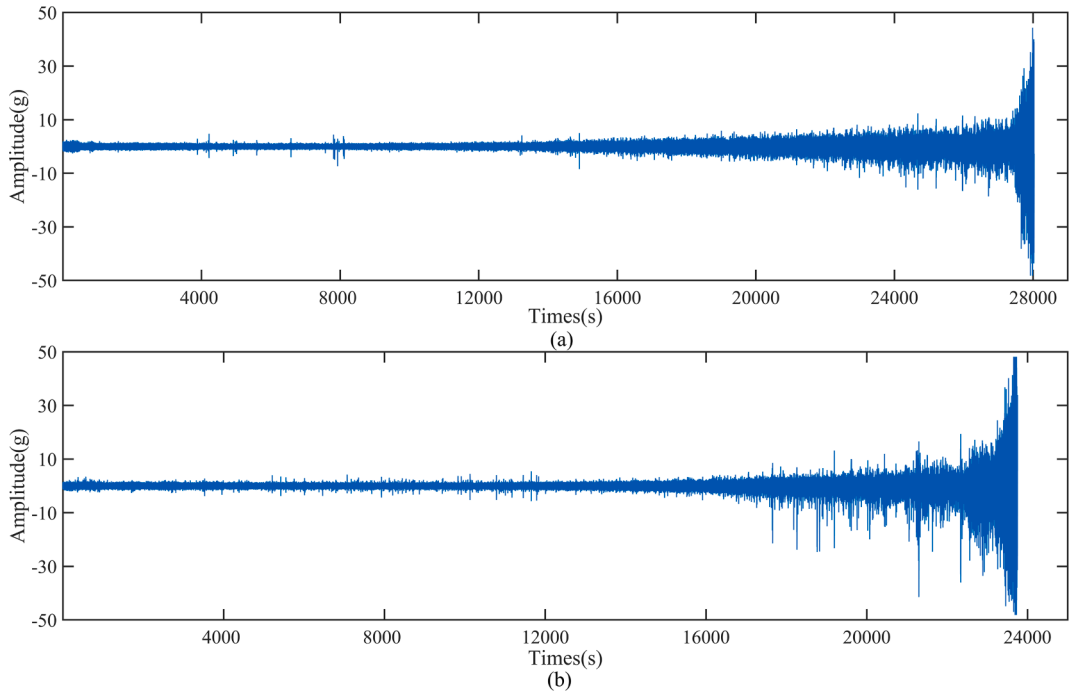


Fig. 5. The life-cycle waveforms of two bearings. (a) bearing1_1, (b) bearing1_3.

input HI series based on$\boldsymbol{\Omega}_{t-1}$, and its update rule is similar to that of the traditional GRU. From the above analysis, we can know that the hidden state of RMGRU contains more trend information than other classical RNN networks. Therefore, the proposed RMGRU has a strong predictive ability, especially in the case of limited samples.

At last, the output hidden state of RMGRU is fed into the fully connected layer for getting the HI at time t, which is written as.

$$\boldsymbol{o}_t = f(\boldsymbol{A}_t \boldsymbol{h}_t + \boldsymbol{d}_t) \tag{14}$$

where $\boldsymbol{o}_t$ denotes the output vector of RMGRU, that is, the predicted HI vector at time t; and f is a linear activate function. Hence, the loss function of DTGRU network is defined as:

$$L(\boldsymbol{o}_t, \widehat{\boldsymbol{o}}_t) = \frac{1}{N}\sum_{n=1}^{N}\left(o_t^{(n)} - \widehat{o}_t^{(n)}\right)^2 \tag{15}$$

where $\widehat{\boldsymbol{o}}_t$ represents the output target of RMGRU, that is, the actual HI vector at time t; and N is the length of output vector.

2.2.4. The RUL prediction approach

With the proposed KLD-based HI and RMGRU, the bearing RUL can be predicted. First, the known HI points are employed to train the RMGRU network, then the unknown HI points are predicted by the trained RMGRU network in sequence (point-by-point) until the predicted HI value exceeds the preset failure threshold. Then, the RUL is calculated via multiplying the number of the predicted points by the recording interval T_s. The specific steps of the proposed RUL prediction approach are given below:

Step1: With the sampling time T and recording interval T_s, the raw vibration signals of bearings are collected. Assume that n samples are acquired in the life cycle of a testing bearing.

Step2: The KLD-based HI construction approach is employed to generate the HI vector from the acquired samples. And the obtained KLD-based HI vector is represented as$\boldsymbol{A} = [a_1, a_2, \cdots, a_n]^T$. The first m elements of $\mathbf{A}$ are utilized as training samples.$\boldsymbol{X}_{train}$, and the rest elements of $\mathbf{A}$ are treated as the testing samples$\boldsymbol{X}_{test}$, i.e. $\boldsymbol{X}_{train} = [a_1, a_2, \cdots, a_m]^T$ and$\boldsymbol{X}_{test} = [a_{m+1}, a_{m+2}, \cdots, a_n]^T$. And the training HI vector $\boldsymbol{X}_{train}$ can be regarded as the labels of RMGRU.

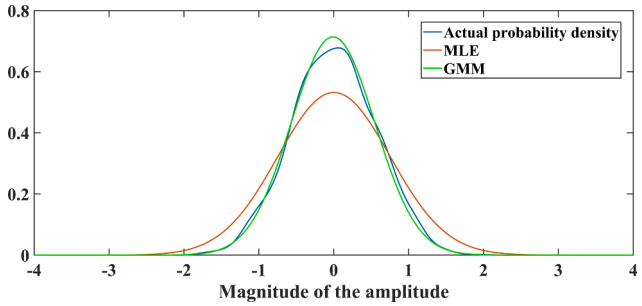


Fig. 6. The probability density functions of the first sample in bearing1_1 estimated by MLE and GMM.

Step3: According to the training samples $\boldsymbol{X}_{train}$ and the number of input neurons k, the input matrix $\mathbf{Y}$ is constructed as.

$$\boldsymbol{Y} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{m-k} \\ a_2 & a_3 & \cdots & a_{m-k+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k+1} & a_{k+2} & \cdots & a_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_{k+1} \end{bmatrix} \tag{16}$$

where.

$$\boldsymbol{Y}_i = [a_i \quad a_{i+1} \quad \cdots \quad a_{m-k+i+1}] \tag{17}$$

Step4: In the training process of RMGRU, the first k row vectors of $\mathbf{Y}$ are used as the input data, the last row vector is treated as the output target. Obviously, the RMGRU network can be viewed as a mapping function g. Therefore, the loss function Eq (15) for updating the parameters $\mathbf{A}$ and $\mathbf{d}$ can be rewritten as:

$$L(\boldsymbol{A}, \boldsymbol{d}) = [g(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_k) - \boldsymbol{Y}_{k+1}]^2 \tag{18}$$

Step5: After training the RMGRU network, the next HI vector can be calculated by inputting the last k raw vectors into the trained RMGRU, and the $(m+1)^{\text{th}}$ HI point is predicted. Repeating this step, the predictive process of HI is described as follows:

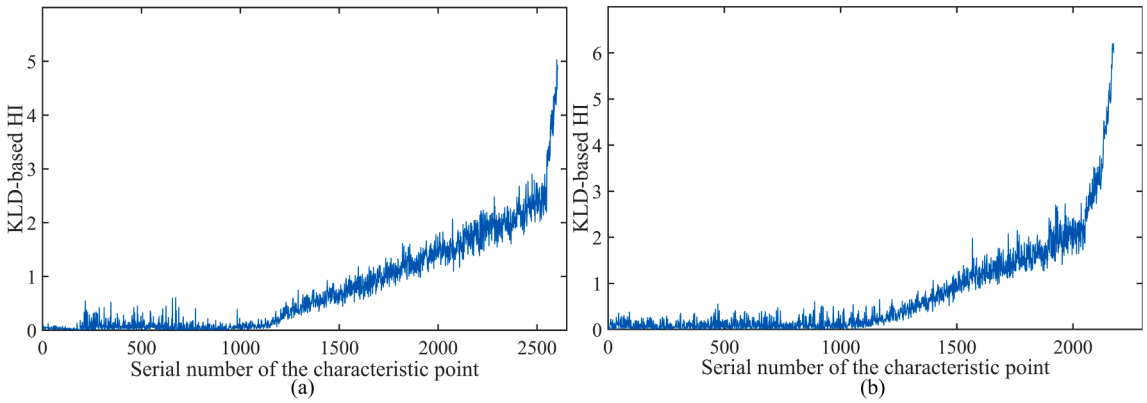


Fig. 7. The KLD-based HIs of two bearings: (a) bearing1_1, (b) bearing1_3.

$$\begin{aligned}\overline{\boldsymbol{Y}}_{k+2} &= g(\boldsymbol{Y}_2, \boldsymbol{Y}_3, \cdots, \boldsymbol{Y}_{k+1})\\ \overline{\boldsymbol{Y}}_{k+3} &= g(\boldsymbol{Y}_3, \boldsymbol{Y}_4, \cdots, \overline{\boldsymbol{Y}}_{k+2})\\ &\vdots\\ \overline{\boldsymbol{Y}}_i &= g(\boldsymbol{Y}_{i-k}, \cdots, \boldsymbol{Y}_{k+1}, \overline{\boldsymbol{Y}}_{k+2}, \cdots, \overline{\boldsymbol{Y}}_{k+(i-k-2)}, \overline{\boldsymbol{Y}}_{k+(i-k-1)})\end{aligned} \tag{19}$$

Step6: When the value of the predicted HI point exceeds the predetermined failure threshold, the bearing RUL can be immediately computed via multiplying the number of the predicted points by the recording interval T_s. By comparing the predicted RUL with the real RUL, the feasibility of the proposed methodology can be confirmed.

3. Analysis of experiment

3.1. Data description

The validation dataset comes from the IEEE 2012 PHM Challenge collected on the PRONOSTIA platform. The test rig, as illustrated in Fig. 4, mainly consists of a drive part, bearing platform, loading device and signal acquisition system, which can implement the bearing performance degradation test and provide the life-cycle datasets. More details about this dataset can be seen in Ref. [43]. In this work, the life-cycle datasets of bearing1_1 and bearing1_3 are utilized to verify the proposed RUL prediction framework, as they have a long and slow degradation process which is helpful for checking the predictive ability of networks. The rotating speed is set as 1800 rpm, and the radial force is set as 4000 N. The sampling frequency is 25.6 kHz, the length of sampling time is 0.1 s, and the sampling is repeated every 10 s (i.e. $T_s = 10$ s). The sample sizes of bearing 1 and bearing 3 are 2803 and 2375, respectively. And the life-cycle waveforms of bearing1_1 and bearing1_3 datasets are respectively drawn in Fig. 5(a) and (b).

3.2. KLD-based HI construction

With regard to the proposed HI construction method, the more accurate the estimated distribution parameters are, the better the degradation process of bearings can be reflected by the KLD-based HI. Compared with the traditional maximum likelihood estimation (MLE) method, GMM can accurately estimate the distribution parameters of raw vibration samples. For instance, the distribution parameters of the first sample in bearing1_1 are estimated by MLE and GMM, and the probability density function is respectively visualized in Fig. 6. From this figure, it can be noted that the distribution parameters estimated by GMM are more accurate than MLE. Therefore, the distribution parameters estimated by GMM are more suitable for constructing the KLD-based HI.

For bearing1_1 and bearing1_3, H is set to 200, that is, the first 200 samples are used to calculate the baseline distribution p_0. Next, the KLD-based HI vector of bearing1_1 and bearing1_3 are constructed, which are respectively visualized in Fig. 7(a) and (b). This figure indicates that the generated HIs are able to well reflect the health stage and degraded trend for these two bearings. When the bearing is at the health stage, the value of KLD-based HI is very close to 0 and fluctuates slightly. The first degradation point is determined by the principle of triple standard deviation, that is, when the KLD-based HI value at one point exceeds the triple standard deviation of the healthy signal, this point is defined as the first degradation point. Then with the degradation of bearings, both the value and fluctuation of KLD-based HI gradually increase. Finally, the value of HI rises dramatically, which indicates that the bearing begins to lose efficacy. Comparing Fig. 5 with Fig. 7, we can see that the variation tendencies of both bearings' vibration amplitudes are similar to the degradation process of the constructed KLD-based HI. No matter for the

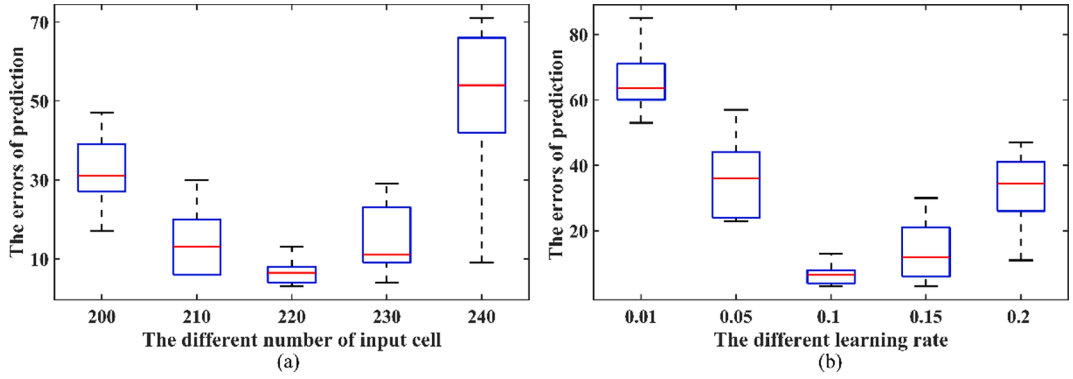


Fig. 8. The optimization of hyper-parameters: (a) the optimization of input cell, (b) the optimization of learning rate.

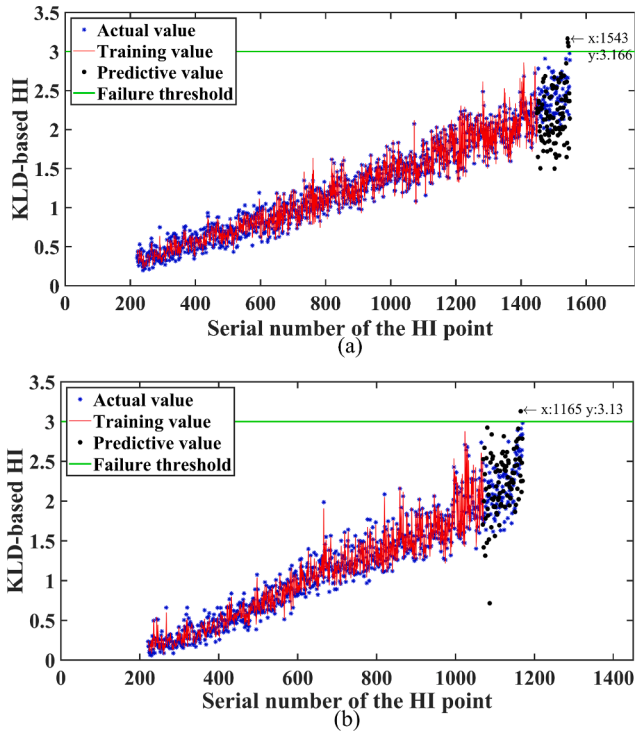


Fig. 9. Predictive results of 100 HI points: (a) bearing1_1, (b) bearing1_3.

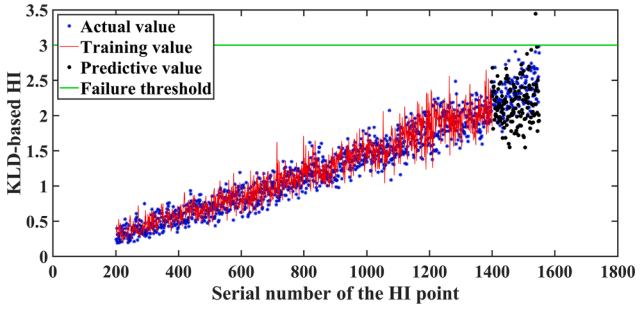


Fig. 10. The prediction of 150 HI points for bearing1_1.

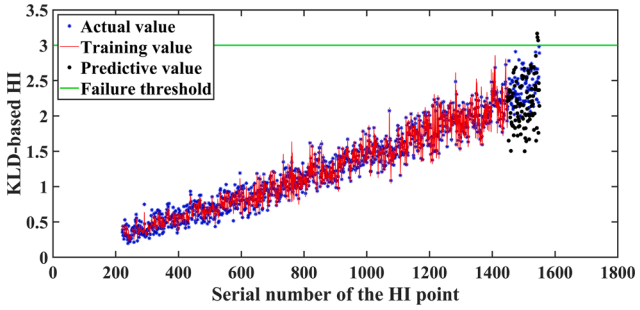


Fig. 11. The prediction of 100 HI points for bearing1_1.

vibration amplitude or the KLD-based HI, the value first stays steady, then gradually increases, finally rises sharply when the bearing fails. Therefore, the constructed KLD-based HI can represent the real health condition of the bearings. Via a series of experiments, we find that the initial bearing failure occurs when the HI value reaches 3, hence it is set to the failure threshold.

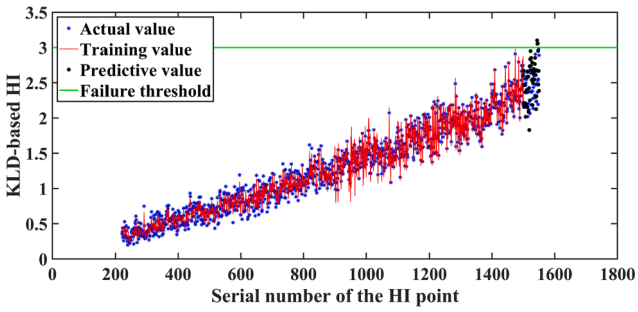


Fig. 12. The prediction of 50 HI points for bearing1_1.

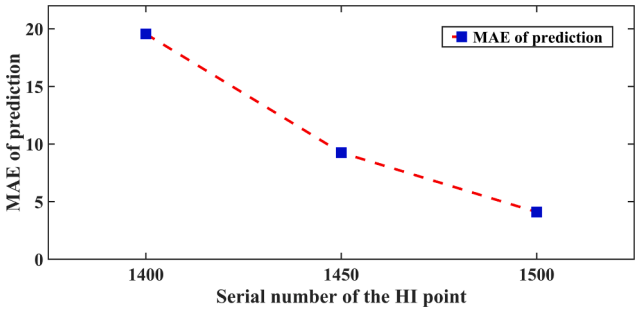
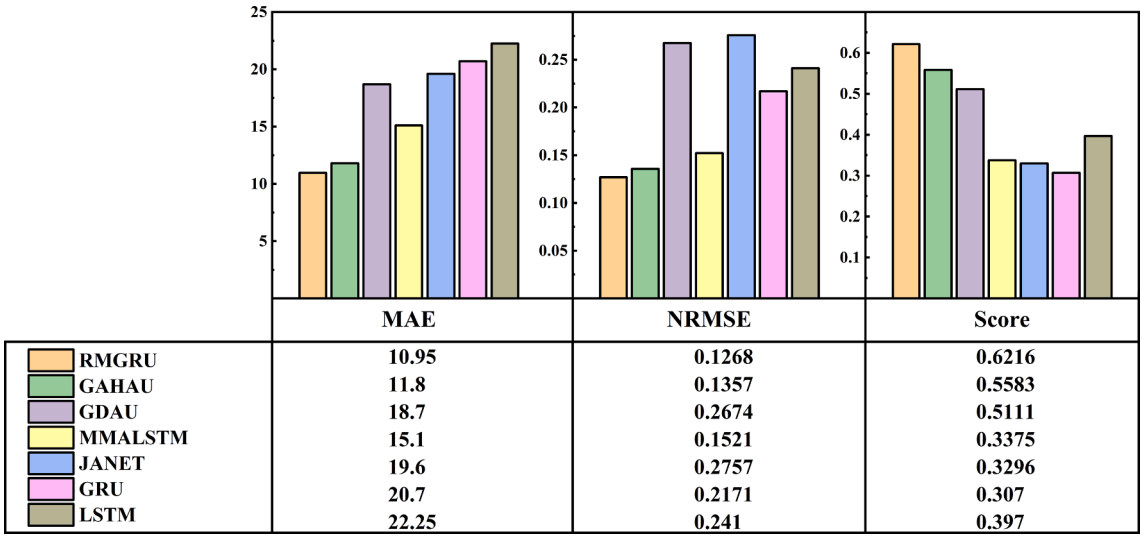


Fig. 13. MAEs of bearing1_1 obtained under different known HI points.

3.3. RUL prediction by RMGRU

Via the constructed KLD-based HI and the proposed RMGRU network, the bearing RUL is predicted. The proposed RUL prediction method pays more attention to the degradation process between the first degradation point and initial failure point, because it contains most of the degradation information. However, in order to fully learn the degradation information, some HI points at the health stage also need to be fed into the prediction model. For the bearing1_1, 1550 points from the 1001st element to the 2550th element are selected, and the first 1450 points are used to predict the variation trend of the last 100 points. Similarly, for the bearing1_3, a total of 1170 points from the 901st element to the 2070th element are selected, and the last 100 points are predicted by the first 1070 points. The initial parameters of RMGRU are generated randomly, and the optimization method adopts the well-known gradient backpropagation algorithm. The RMGRU network framework consists of an input layer, a hidden layer and an output layer. Suppose that the numbers of input cells and output cells are respectively N_i and 1. The number of hidden cells is calculated by $\sqrt{N_i+1}+20$, then it is rounded. And the number of input cells and the learning rate are the main hyper-parameters to affect the RMGRU model. In this paper, the grid search technology is employed to optimize these two hyper-parameters. The errors of bearing1_1 RUL prediction results obtained by different input cells and learning rates are demonstrated in Fig. 8. This figure indicates that the error of RUL prediction decreases first and then increases with the increase of input cells' number and learning rates. It can be known from Fig. 8 that the optimal number of input cells is 220 and the optimal learning rate is 0.1.

With the proposed RUL prediction approach and optimal hyper-parameters, the predictive results of bearing1_1 and bearing1_3 are obtained. The training, predictive and actual RUL curves of bearing1_1 are simultaneously drawn in Fig. 9(a), while those of bearing1_3 are simultaneously drawn in Fig. 9(b). Fig. 9 shows that the training and predictive RULs are very close to the real ones at almost all time instants, and the degradation trends of training and predictive HI curves are also approximate to that of the actual HI curve. In Fig. 9(a), the number of HI points that don't exceed the preset failure threshold is 93, while it is 95



	MAE	NRMSE	Score
RMGRU	10.95	0.1268	0.6216
GAHAU	11.8	0.1357	0.5583
GDAU	18.7	0.2674	0.5111
MMALSTM	15.1	0.1521	0.3375
JANET	19.6	0.2757	0.3296
GRU	20.7	0.2171	0.307
LSTM	22.25	0.241	0.397

Fig. 14. The evaluation indexes of different prediction networks for bearing1_1.

in Fig. 9(b). It then follows that the RULs of bearing1_1 and bearing1_3 are respectively 930 and 950 s.

For thoroughly displaying the long short-term predictive ability of RMGRU, the last 150, 100 and 50 unknown HI points of bearing1_1 are respectively predicted. Then the predictive results of HI for these three cases are respectively depicted in Figs. 10-12. These figures indicate that the HI values of training and predicting are approximate to those of the actual HI values. Even if the number of predicted points reaches 150, the predictive degradation tendency still accords with the actual one. Meanwhile, we can easily know that more HI points should be if we address to realize the more accurate prediction. Hence intuitively speaking, the proposed RMGRU network has strong robustness for the bearing RUL prediction. Owing to the randomness of parameter initialization, the mean absolute error (MAE) is selected as the evaluation metric of RUL prediction for enhancing the confidence of the predicting results. In each case, 20 parallel experiments are carried out, then MAE is calculated by.

$$MAE = \frac{1}{20}\sum_{i=1}^{20}|TRul - PRul_i| \tag{20}$$

where $TRul$ represents the actual RUL of bearing, $PRul_i$ denotes the predicted RUL of the ith experiment. The obtained MAEs of three cases are drawn in Fig. 13. This figure shows the value of MAE gradually decreases with the increase of the known HI points. This is mainly because the RMGRU network can dig out more useful degradation information from more known HI points. Moreover, based on these 20 experiments, the predictive performance is evaluated by the average predictive accuracy (APA). The formula of APA is given by:

$$APA = 1 - \frac{1}{20}\sum_{i=1}^{20}\frac{|TRul - PRul_i|}{TRul} \times 100\% \tag{21}$$

The APA is calculated as 91.8%, when predicting 50 HI points, that is, the actual RUL is 500 s. Similarly, the APA is 86.8% for the prediction of 150 HI points, while that can reach 90.7% for the prediction of 100 HI points. The calculation results of APA further show that the proposed RMGRU is able to well deal with the long short-term RUL prediction.

3.4. Comparison with other prediction model

RMGRU is actually an improved RNN. To verify the superiority of the RMGRU, the traditional RNN networks including LSTM and GRU, and the state-of-the-art RNN-based networks including gated dual attention unit (GDAU) [12], macroscopic-microscopic attention in LSTM (MMALSTM) [16], just another network (JANET) [44] and gated adaptive hierarchical attention unit (GAHAU) [17] are employed for comparison. Compared with the proposed RMGRU network, these existing predictive networks don't make full utilization of the historical state information to reduce the forgetting of trend information. In the contrast experiment, the last 100 elements of the KLD-based HI vector are predicted. For the evaluation of the predictive performance from multiple perspectives, several quantitative evaluation indexes, such as MAE, normalized root mean square error (NRMSE), and a score function [40], are employed. MAE and NRMSE are respectively applied to measure the absolute error and relative error of the RUL prediction results. The score function is used to evaluate the overestimation or underestimation result, which can assign a larger decrement to the overestimated result than the underestimated result. This is because that the overestimation of RUL may result in more serious losses than its underestimation. NRMSE and Score are respectively formulated as follows:

$$NRMSE = \sqrt{\frac{1}{20}\sum_{i=1}^{20}(TRul - PRul_i)^2}\Big/\frac{1}{20}\sum_{i=1}^{20}PRul_i \tag{22}$$

$$Score = \frac{1}{20}\sum_{i=1}^{20}Y_i \tag{23}$$

where Y_i is defined as:

$$Y_i = \begin{cases} \exp(-\ln(0.5) \times (AR_i/5)), & AR_i \leqslant 0 \\ \exp(\ln(0.5) \times (AR_i/20)), & AR_i > 0 \end{cases} \tag{24}$$

$$AR_i = \frac{TRul - PRul_i}{TRul} \times 100 \tag{25}$$

For a fair comparison, all the comparative networks are optimized. With the RULs obtained by all the predictive networks, the three evaluation indexes are respectively calculated and illustrated in Fig. 14. It is obvious that the MAE and NRMSE of RMGRU are the smallest while the Score of RMGRU is the highest, thus the RMGRU network has better predictive performance than the classical predictive networks. The reason is that the proposed RMGRU is able to relearn the oblivious trend characteristics in time, namely, RMGRU can remember more useful information of degradation trend than the current RNN-based networks. The results of contrast experiments also indicate that the reuse of historical state information is more effective than just improving the cell structure.

4. Conclusions

In order to decrease the prediction error and enhance the robustness of the long-term prediction under the limited and unlabeled samples, this article researches a novel RUL prediction approach consisting of KLD-based HI and RMGRU. In terms of HI generation, the unsupervised KLD-based HI with a distinct degradation trend is first extracted by GMM and KL divergence from the original bearing vibration signals. With regard to RUL prediction, RMGRU is innovatively proposed by relearning the amnesic knowledge of degradation trend, and it can effectively predict the tendency of the KLD-based HI vector. Then via a failure threshold, the bearing RUL can be estimated. The proposed RUL prediction approach has been successfully verified by the IEEE PHM 2012 bearing datasets, and it can be applied to the long short-term prediction with high precision. Furthermore, the results of contrast show that RMGRU has a stronger predictive ability than GAHAU, GDAU, MMALSTM, JANET, GRU and LSTM. Therefore, the proposed RUL prediction framework is more conducive to the prediction of bearing RUL.

In practice, the difficulty of determining the failure threshold under different working conditions limits the application of the proposed method. Therefore, constructing a HI with a uniform failure threshold under different working conditions will be explored in the future study. Moreover, it should be noted that the proposed RMGRU is mainly applicable to the prediction of the long degradation process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work described in this paper was supported by the National Natural Science Foundation of China (nos. 52175075 and 62033001), Chongqing Research Program of Basic Research and Frontier Exploration (no. cstc2021ycjh-bgzxm0157), and National High Tech ship research project (No.360 [2019] issued by MIIT, China).

References

[1] Z. Liu, H. Liu, H. Wang, W. Jia, D. Zhang, J. Tan, A multi-head neural network with unsymmetrical constraints for remaining useful life prediction, Adv. Eng. Inf. 50 (2021), 101396.
[2] Y. Qin, X. Wu, J. Luo, Data-model combined driven digital twin of life-cycle rolling bearing, IEEE Trans. Ind. Inf. 18 (3) (2022) 1530–1540.
[3] X. Li, W. Zhang, H. Ma, Z. Luo, X. Li, Degradation Alignment in Remaining Useful Life Prediction Using Deep Cycle-Consistent Learning, IEEE Trans. Neural Netw. Learn. Syst., in press (2021).
[4] W. Yu, I.I.Y. Kim, C. Mechefske, Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme, Mech. Syst. Sig. Process. 129 (2019) 764–780.
[5] L. Cui, X. Wang, H. Wang, J. Ma, Research on Remaining Useful Life Prediction of Rolling Element Bearings Based on Time-Varying Kalman Filter, IEEE Trans. Instrum. Meas. 69 (2020) 2858–2867.
[6] Y. Hu, P. Baraldi, F. Di Maio, E. Zio, Online Performance Assessment Methodfor a Model-Based Prognostic Approach, IEEE Trans. Reliab. 65 (2016) 718–735.
[7] Y. Cheng, K. Hu, J. Wu, H. Zhu, X. Shao, A convolutional neural network based degradation indicator construction and health prognosis using bidirectional long short-term memory network for rolling bearings, Adv. Eng. Inf. 48 (2021), 101247.
[8] X. Xin, Y. Tu, Y. Chen, V. Stojanovic, H. Wang, K. Shi, S. He, T. Pan, Online reinforcement learning multiplayer non-zero sum games of continuous-time Markov jump linear systems, Appl. Math. Comput. 412 (2022), 126537.
[9] O.O. Aremu, D. Hyland-Wood, P.R. McAree, A Relative Entropy Weibull-SAX framework for health indices construction and health stage division in degradation modeling of multivariate time series asset data, Adv. Eng. Inf. 40 (2019) 121–134.
[10] W. Ahmad, S.A. Khan, J.-M. Kim, A Hybrid Prognostics Technique for Rolling Element Bearings Using Adaptive Predictive Models, IEEE Trans. Ind. Electron. 65 (2018) 1577–1584.
[11] Z. Shi, A. Chehade, A dual-LSTM framework combining change point detection and remaining useful life prediction, Reliab. Eng. Syst. Saf. 205 (2021), 107257.
[12] Y. Qin, D. Chen, S. Xiang, C. Zhu, Gated dual attention unit neural networks for remaining useful life prediction of rolling bearings, IEEE Trans. Ind. Inf. 17 (9) (2021) 6438–6447.
[13] K. Deng, X. Zhang, Y. Cheng, Z. Zheng, F. Jiang, W. Liu, J. Peng, A remaining useful life prediction method with long-short term feature processing for aircraft engines, Appl. Soft Comput. 93 (2020), 106344.
[14] L. Ren, X. Cheng, X. Wang, J. Cui, L. Zhang, Multi-scale Dense Gate Recurrent Unit Networks for bearing remaining useful life prediction, Future Generation Computer Systems 94 (2019) 601–609.
[15] J. Zhu, N. Chen, C. Shen, A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions, Mech. Syst. Sig. Process. 139 (2020), 106602.
[16] Y. Qin, S. Xiang, Y. Chai, H. Chen, Macroscopic-Microscopic Attention in LSTM Networks Based on Fusion Features for Gear Remaining Life Prediction, IEEE Trans. Ind. Electron. 67 (12) (2020) 10865–10875.
[17] D. Chen, Y. Qin, J. Luo and S. Xiang, Gated adaptive hierarchical attention unit neural networks for the life prediction of servo motors, IEEE Transactions on Industrial Electronics, https://doi.org/10.1109/TIE.2021.3112987, in press.
[18] M.M. Islam, A.E. Prosvirin, J. Kim, Data-driven prognostic scheme for rolling-element bearings using a new health index and variants of least-square support vector machines, Mech. Syst. Sig. Process. 160 (2021), 107853.
[19] L.M. Carvalho; J. Teixeira; M. Matos, Modeling wind power uncertainty in the long-term operational reserve adequacy assessment: A comparative analysis between the Naïve and the ARIMA forecasting models, 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2016.
[20] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780.
[21] K. Cho, B.V. Merrienboer, C. Gulcehre, D. Ba Hdanau, F. Bougares, H. Schwenk, Y. J.C.S. Bengio, Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, arXiv:1406.1078, (2014).
[22] S. Xiang, Y. Qin, J. Luo, H. Pu, B. Tang, Multicellular LSTM-based deep learning model for aero-engine remaining useful life prediction, Reliab. Eng. Syst. Saf. 216 (2021), 107927.
[23] C.-G. Huang, H.-Z. Huang, Y.-F. Li, A Bidirectional LSTM Prognostics Method Under Multiple Operational Conditions, IEEE Trans. Ind. Electron. 66 (2019) 8792–8802.
[24] Y. Qin, J. Zhou, D. Chen, Unsupervised health Indicator construction by a novel degradation-trend-constrained variational autoencoder and its applications, IEEE/ASME Trans. Mechatron. 27 (3) (2022) 1447–1456.
[25] D. She, M. Jia, A BiGRU method for remaining useful life prediction of machinery, Measurement 167 (2021), 108277.
[26] T. Bluche, J. Louradour, R. Messina, Scan, Attend and Read: End-to-End Handwritten Paragraph Recognition with MDLSTM Attention, 2017 14th IAPR international conference on document analysis and recognition (ICDAR), IEEE, 2017.
[27] H. Zheng, J. Fu, M. Tao, J. Luo, Learning Multi-attention Convolutional Neural Network for Fine-Grained Image Recognition, 2017 IEEE International Conference on Computer Vision (ICCV), 2017.
[28] Y. Chen, G. Peng, Z. Zhu, S. Li, A novel deep learning method based on attention mechanism for bearing remaining useful life prediction, Appl. Soft Comput. 86 (2020), 105919.
[29] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, X. Li, Machine Remaining Useful Life Prediction via an Attention-Based Deep Learning Approach, IEEE Trans. Ind. Electron. 68 (2021) 2521–2531.
[30] B. Li, B. Tang, L. Deng, M. Zhao, Self-Attention ConvLSTM and Its Application in RUL Prediction of Rolling Bearings, IEEE Trans. Instrum. Meas. 70 (2021) 1–11.
[31] M. Ma, Z. Mao, Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction, IEEE Trans. Ind. Inf. 17 (2021) 1658–1667.
[32] L. Zhang, J. Wen, A systematic feature selection procedure for short-term data-driven building energy forecasting model development, Energy Build. 183 (2019) 428–442.
[33] Z. Huang, Z. Xu, X. Ke, W. Wang, Y. Sun, Remaining useful life prediction for an adaptive skew-Wiener process model, Mech. Syst. Sig. Process. 87 (2017) 294–306.
[34] M. Zhao, S. Zhong, X. Fu, et al., Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis, IEEE Trans. Ind. Electron. 68 (3) (2021) 2587–2597.
[35] L. Guo, Y. Lei, N. Li, T. Yan, N. Li, Machinery health indicator construction based on convolutional neural networks considering trend burr, Neurocomputing 292 (2018) 142–150.
[36] P. shankar kumar, L.A. Kumaraswamidhas, S.K. Laha, Bearing degradation assessment and remaining useful life estimation based on Kullback-Leibler divergence and Gaussian processes regression, Measurement 174 (2021), 108948.
[37] D. Reynolds, Gaussian mixture models, Encycl. Biometric Recognit. 31 (2008) 1047–1064.

[38] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, J. Roy. Stat. Soc. 39 (1997) 1–38.
[39] S. Kullback, R. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1951) 79–86.
[40] A. Rai, S.H. Upadhyay, An integrated approach to bearing prognostics based on EEMD-multi feature extraction, Gaussian mixture models and Jensen-Rényi divergence, Appl. Soft Comput. 71 (2018) 36–50.
[41] V. Mnih, N. Heess, A. Graves, K. Kavukcuoglu, Recurrent Models of Visual Attention, Proc. 27th Int Conf. Neural Inf. Process. Syst. (2014) 2204–2212.
[42] H. Ebbinghaus, Memory: A Contribution to Experimental Psychology, Annals of Neurosciences 20 (4) (2013) 155–156.
[43] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Morello, N. Zerhouni, C. Varnier, PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests, IEEE Int. Conf. Prognost. Health Manage. (2012).
[44] V. Jos, J. Lasenby, The unreasonable effectiveness of the forget gate, arXiv: 1804.04849v3 (2018).