IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 19, 2020, accepted June 15, 2020, date of publication June 26, 2020, date of current version July 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005165

A New State Recognition and Prognosis Method Based on a Sparse Representation Feature and the Hidden Semi-Markov Model

YUN-FEI MA, XISHENG JIA, QIWEI HU, HUAJUN BAI, CHIMING GUO, AND SHUANGCHUAN WANG

Department of Equipment Command and Management, Army Engineering University, Shijiazhuang 050003, China

Corresponding author: Yun-Fei Ma (fcz1992@sina.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71871220.

ABSTRACT Equipment degradation state recognition and prognosis are considered two significant parts of a prognostics and health management (PHM) system that help to reduce downtime and decrease economic losses. In this paper, a sparse representation (SR) feature is proposed as a new degradation feature, and the hidden semi-Markov model (HSMM) is established. The new method offers three significant advantages over the traditional HSMM. (1) Since the degradation information is incomplete, a Gaussian mixture model (GMM) is used here for degradation data clustering and state division. (2) A new degradation feature based on the wavelet packet transform (WPT) and SR can better extract the structural information of the collected signal and reflect the degradation characteristics. (3) To conduct remaining useful life (RUL) predictions, an improved model is proposed, which adds a control variable that can dynamically adjust the state duration. The effectiveness of the proposed method is demonstrated using 8 groups of bearing data from the Center for Intelligent Maintenance Systems (IMS). The results show that the HSMM with the SR feature achieves the best recognition accuracy, of 85.28%. Moreover, the improved prediction model achieves a prediction accuracy of 86.11% on average for 8 bearings.

INDEX TERMS Sparse representation, hidden semi-Markov model, degradation recognition, feature extraction, prognostic, bearing.

NOMENCLATURE

PHM	prognostics and health management
RUL	remaining useful life
HMM	hidden Markov model
HSMM	hidden semi-Markov model
HOPF	high-order particle filter
EEMD	ensemble empirical mode decomposition
AHSMM	adaptive hidden semi-Markov model
GMM	Gaussian mixture model
WPT	wavelet packet transform
SR	sparse representation
EM	expectation and maximization
K-SVD	k-means singular value decomposition
CS	compressive sensing
BCS	Bayesian compressive sensing
PC	the Partition Coefficient
SC	the Partition Index
SI	the Separation Index
XB	the Xie and Beni's Index
IMS	intelligent maintenance systems
PCA	Principal component analysis
WPE	wavelet packet energy
MC	Monte Carlo
K	the number of macro-states
s_t	the segment at time t
q_i	the end-point of macro-state x_i
$X = \{x_1, x_2, \ldots, x_K\}$	the macro-states sequence
π	initial state distribution
A	the state transition probability distribution
$Y = \{y_1, y_2, \ldots, y_L\}$	the observation model
$O = \{o_1, o_2, \ldots, o_T\}$	the observation sequence
B	the observation probability distribution
H	the HSMM model
$\boldsymbol{y}$	the signal to be decomposed
$\boldsymbol{D}$	the redundant dictionary
m	the dimension of signal $\boldsymbol{y}$

The associate editor coordinating the review of this manuscript and approving it for publication was Nagarajan Raghavan.

n	the number of atoms in dictionary $\boldsymbol{D}$
$\{\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_n}\}$	the atoms in dictionary $\boldsymbol{D}$
$\boldsymbol{x}$	the sparse coefficients
$\boldsymbol{Y}$	the training matrix in dictionary learning
$\boldsymbol{X}$	the set of sparse coefficients
$\boldsymbol{x_i}$	the i-th row of matrix $\boldsymbol{X}$
$\boldsymbol{E_i}$	the residual matrix in the K-SVD learning algorithm
$\boldsymbol{\omega_i}$	a nonzero index set of $\boldsymbol{x_i}$
$\boldsymbol{\Omega_i}$	the matrix replacing the elements in set $\boldsymbol{\omega_i}$ with element 1
$\boldsymbol{E_R^i}$	the product of $\boldsymbol{E_i}$ and $\boldsymbol{\Omega_i}$
$\boldsymbol{x_R^i}$	the product of $\boldsymbol{x_i}$ and $\boldsymbol{\Omega_i}$
$\boldsymbol{U}$, $\boldsymbol{\Delta}$, $\boldsymbol{V}$	the SVD decomposition results of matrix $\boldsymbol{E_R^i}$
$\boldsymbol{u_1}$, $\boldsymbol{v_1}$	the first column of matrices $\boldsymbol{U}$ and $\boldsymbol{V}$
$\boldsymbol{wp_i}$	the i-th subband of wavelet packet decomposition
$\boldsymbol{\Phi_i}$	the measurement matrix of subband i
$\boldsymbol{\lambda_i}$	the measurement vector of subband i
$\boldsymbol{\theta_i}$	the sparse vector corresponding to dictionary $\boldsymbol{D_i}$
δ_i	the reconstruction residual of subband i
ϵ	denotes the number of subbands
$\boldsymbol{Di}$	the dictionary trained for subband i
$P(x_l)$	the macro-state duration density
$\mu(x_l)$	the duration mean
$\sigma^2(x_l)$	the duration variance
P_i	the probability of equipment state i
RL_i	the estimated RUL of state i
U_i	the unified variable of the three duration values of state i
a	the number of degradation states
$\overline{U_{i-1}}$	the average duration of the historical data
k_i	is the variable controlling the duration of the degradation state i
F_i	the degradation feature of state i
$F(t)$	represents the degradation feature of time t

I. INTRODUCTION

Equipment monitoring, detection, and management have played important roles in modern militaries. Prognostics and health management (PHM) systems have been applied in both military [1] and civilian [2]–[4] areas. In the military, the efficiency and availability of equipment are increased in many applications where equipment safety and reliability are critical. The PHM process involves observing degradation features, evaluating the current states, and predicting the remaining useful life (RUL). Generally, a prerequisite to the deployment of a PHM is effective diagnostics and prognostics.

Current health diagnostics and prognostics methods can be divided into three kinds: physical models, data-driven models, and mathematical models. The drawbacks of physical models are low accuracy and high cost. The drawback of data-driven models stems from the large amount of data, which consumes much time in practice. Thus, in this paper, a mathematical model called the hidden Markov model (HMM) is introduced for equipment diagnostics and prognostics. Most machinery degradation is a slow process from normal to failure states. Due to influences of the equipment and noise, the relationships between internal states and devices become complicated. Assessing equipment states using external measurements is a double stochastic process, which is consistent with the HMM. This model has a rich mathematical structure, which fits actual equipment degradation well. Moreover, the HMM combines both diagnostics and prognostics in a unified framework, and this can be seen as another advantage.

The HMM is increasingly popular in many fields, such as speech recognition [5], [6] and handwritten word recognition [7]. The hidden semi-Markov model (HSMM), constructed by adding a temporal component to the HMM structure, has satisfactorily addressed the limitations of the HMM for the Markov property and can be directly used in prognostics. In recent years, the HSMM has been widely used in degradation assessment and fault prognostics. Dong *et al.* [8] applied the HSMM to UH-60A Blackhawk planetary carriers and a hydraulic pump. The diagnosis accuracy was increased by 45% and 81% for test cell and on-aircraft measurements, respectively. They assume that the degradation states are preset and use the wavelet coefficients as the degradation features. Xiao *et al.* [9] proposed a modified duration-dependent HSMM for online condition monitoring. This model was defined as duration dependent, which is realistic. Then, a high-order particle filter (HOPF) method was applied to predict the RUL. The hidden states were determined according to experience-based failure knowledge, and the ensemble empirical mode decomposition (EEMD) method was used to extract the fault feature information. Cannarile *et al.* [10] proposed a homogeneous continuous-time finite-state hidden semi-Markov model (HSMM) to estimate the degradation states. The diagnosis effectiveness was thoroughly tested with the bearing data. However, the research on prediction effects is relatively weak, for it has only judged the degradation level of observed signals. Liu *et al.* [11] improved the HSMM by using a sequential Monte Carlo method to describe the probability relationships between degradation states and observations. They also developed a new online health prognostic method for RUL estimation. Furthermore, they proposed an adaptive hidden semi-Markov model (AHSMM) [12] for multisensor equipment diagnosis and prognosis. The wavelet packet energy (WPE) is used as the degradation feature in this article.

From the related work mentioned above, the application of the HSMM faces three problems: (1) Determining the degradation states. Traditional works assume that the

degradation states are known. However, the number of these states may not be accurate if there is a lack of corresponding prior knowledge. For example, if only run-to-failure data can be obtained, it is impossible to directly obtain the degradation states, and this limits the application of the HSMM. Zeng [13] proposed an algorithm based on the minimum description length (MDL) principle. Teng *et al.* [14] proposed a method combining K-means and cross-validation to optimize this number. Zhang *et al.* [15] established a group of evaluating features for choosing appropriate state number. They set a range of available state number and enumerated them in order. The evaluating features were calculated for selecting optimized degradation state number. In practice, it is difficult to verify the efficiency of the state number according to the clustering effects, so we adopt the method in [15] for state number selection. (2) Degradation feature selection. According to the literature review, wavelet packet decomposition is the most commonly used feature extraction technique. Moreover, the reference [16] used a wavelet correlation feature scale entropy as the input of HSMM. The feature based on the information entropy theory is presented to get high signal-to-noise scales wavelet coefficients. Wang *et al.* [17] used the relative entropy fusion for hydraulic pumps fault predicting. A features fusion algorithm was also presented for making full use of degradation feature information. In addition, a new approach based on generalized mathematical morphological particle is proposed in [18]. Although the above features have achieved good results, there is still room for improvement to achieve better recognition and prediction performance. (3) Prognostics using the HSMM. Combined with the HSMM, the particle filter (PF) [9] and Monte Carlo (MC) methods [19] are most commonly used for estimating the RUL. However, these techniques are computationally intense and not suitable for online RUL prediction in real industrial applications. Reference [20] proposed a relatively simple model for remnant life prediction. The RUL of each state is obtained by adding the duration of that state and the subsequent degradation states. In this way, the calculated RUL is larger than the actual one unless the duration of each state is short enough.

This paper improves the traditional HSMM by addressing the above three problems. First, we propose a Gaussian mixture model (GMM) [21] based on the timing factor. The GMM has been demonstrated to be efficient for speech recognition [22]. In addition, the timing factor ensures that the GMM's clustering follows the timing order. Moreover, a new feature based on the wavelet packet transform (WPT) [23] and sparse representation (SR) [24], [25] is proposed in this paper to observe degradation. The wavelet transform is outstanding among equipment condition recognition methods due to its multiresolution ability. SR allows the salient information in a signal to be conveyed with a linear combination of elementary components, which are called atoms. To date, it has been successfully applied to a variety of problems, including compressive sensing (CS) [26]–[28], signal denoising [29], and fault classification [30]. In this study, SR is applied to find concise, high-level representations of WPT subbands that match the structure of the degradation signal by using a learned dictionary. Finally, an improved prediction model based on the HSMM is proposed. To obtain an accurate duration, a control variable k is introduced in the model, and it can be calculated using the degree of degradation. Thus, a more accurate RUL value can be obtained through dynamically adjusting the duration.

This paper is organized as follows: Section 2 provides a general theoretical background. The GMM-SR-HSMM framework for degradation recognition and prognostic is presented in Section 3. Section 4 gives a real application of bearing from IMS center. Finally, Section 5 summarizes this paper and draws conclusions.

II. METHODOLOGY

A. THE HIDDEN SEMI-MARKOV MODEL

The hidden Markov [5] is a double stochastic process. One process is a Markov chain describing the state transitions; this is the basic process in which states cannot be directly observed. The states in this basic process are called the hidden states. Another stochastic process describes the correspondence between observations and hidden states. The HSMM is different from the HMM due to its explicit state duration, and the duration is a random value.

The states in a segmental HSMM are called macro-states and each macro-state consists of several micro-states. Only the transition between macro-states follows Markov principles. Fig. 1 displays a basic framework of the HSMM. An HSMM can be described with following parameters [8].

(1) The number of macro-states is K, the segment at time t is s_t, and the end-point is q_i. All macro-states can be represented as $X = \{x_1, x_2, \ldots, x_K\}$.

(2) The initial state distribution is $\pi = \{\pi_i | \pi_i = P(x_i = i), i = 1, 2 \ldots, K\}$, and the state transition probability distribution is

$$A = \{a_{i,j} \,|a_{i,j} = P(s_t = x_j \,|s_{t-1} = x_i)\} \tag{1}$$

(3) $Y = \{y_1, y_2, \ldots, y_L\}$ is the observation model and $O = \{o_1, o_2, \ldots, o_T\}$ is the observation sequence with T elements. The observation probability distribution in state I, $B = \{b_i(v)\}$, where

$$B = \{b_i(v) \,|b_i(v) = P(o_v \,|s_t = x_i)\} \tag{2}$$

Therefore, a complete HSMM model requires the specifications: A, B, π, K, L.

In the HMM, the duration time is measured using a durational probability density function, and this is a geometrically decaying function. The reference [31] demonstrated that this is a less accurate duration model than that of the HSMM. In real applications, the HSMM describes three basic problems, and the corresponding solutions are given as follows:

(1) Evaluating: Given observation $O(o_1, o_2, \ldots, o_T)$ and HSMM h, the probability of appearance for observation O under h is represented as $P(O \mid h)$. This value can be solved

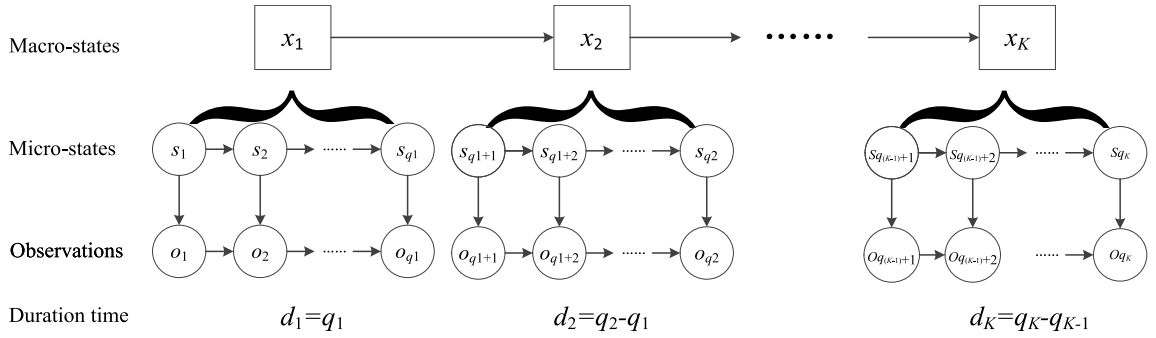


FIGURE 1. The general framework of the HSMM.

by forward–backward algorithm. (2) Decoding: Given observation $O(o_1, o_2, \ldots, o_T)$ and HSMM h, the most possible hidden states sequence of h is denoted as $S(s_1, s_2, \ldots, s_T)$. This sequence, which is also explained as states transition path, can be obtained by Viterbi algorithm. (3) Learning. Given observation $O(o_1, o_2, \ldots, o_T)$ and HSMM h, the most possible parameters set to maximize $P(O \mid h)$ can be denoted as $h = \{\mathrm{A}, \mathrm{B}, \pi, K, L\}$. This set is solved by Baum–Welch algorithm. The process of three algorithms can be found in [32] for more details.

B. SPARSE LEARNING

The SR theory [24], [25] has achieved great popularity over the years. The aim of the SR is to find a linear combination of relatively small atoms to represent the original signal.

The basic idea of SR is simple: Given a signal $\boldsymbol{y} \in R^m$, a redundant dictionary $\boldsymbol{D} = \{\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_n}\}$ is established that consists of n atoms. The aim of SR is to find a number of atomic $\boldsymbol{d_1}, \boldsymbol{d_2}, \ldots, \boldsymbol{d_n}$, such that original signal $\boldsymbol{y}$ can be approximated sparsely by

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} = \sum_{j=1}^{n} \boldsymbol{d_j} x_j \tag{3}$$

If $m < n, \boldsymbol{D}$ is termed as an overcomplete dictionary. Given the signal $\boldsymbol{y}$, finding its representation is done by the following sparse approximation problem:

$$\arg\min \|\boldsymbol{x}\|_0 \quad \text{s.t. } \boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} \tag{4}$$

where the $\|\boldsymbol{x}\|_0$ is a count of the number of non-zeroes in the vector $\boldsymbol{x}$. And the process of solving above optimization problem can be transformed to

$$\arg\min \|\boldsymbol{x}\|_1 \quad \text{s.t. } \boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} \tag{5}$$

The dictionary $\boldsymbol{D}$ can either be predefined or learned. The fixed dictionary is usually built through the following framework: the Fourier, Wavelet, and discrete cosine transforms. This kind of dictionary has a strong dependence on prior knowledge. Conversely, learned dictionaries, which deduce the dictionary from training data, can directly capture the specific features in the original signal. For example, Aharon *et al.* [33] proposed the K-SVD algorithm. It involves a sparse coding stage, which is based on a pursuit method, followed by an update step.

In K-SVD learning, we assumes that the training matrix is $\boldsymbol{Y} = [\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_l}]$ and sparse matrix is $\boldsymbol{X}$. The process of learning can be described by (6) as follows:

$$\min_{\boldsymbol{D}} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2 \quad \text{s.t. } \|\boldsymbol{x_i}\|_0 \le T \tag{6}$$

where $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_l}]$ represents the sparse coefficients. The residual matrix $\boldsymbol{E_i}$ is computed by (7), where $\boldsymbol{d_j}$ represents the j-th column and $\boldsymbol{x_i}$ represents the i-th row.

$$\begin{aligned} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2 &= \left\| \boldsymbol{Y} - \sum_{j=1}^{m} \boldsymbol{d_j}\boldsymbol{x_j} \right\|_F^2 \\ &= \left\| (\boldsymbol{Y} - \sum_{j \ne i}^{m} \boldsymbol{d_j}\boldsymbol{x_j}) - \boldsymbol{d_i}\boldsymbol{x_i} \right\|_F^2 = \|\boldsymbol{E_i} - \boldsymbol{d_i}\boldsymbol{x_i}\|_F^2 \end{aligned} \tag{7}$$

Since most of the elements in $\boldsymbol{x_i}$ are zero, the residual matrix $\boldsymbol{E_i}$ can be restricted through choosing nonzero elements. We define set $\boldsymbol{\omega_i}$ as follow

$$\boldsymbol{\omega_i} = \{v \mid \boldsymbol{x_i}(v) \ne 0\} \tag{8}$$

Then, we substitute the elements in set $\boldsymbol{\omega_i}$ with element 1 and obtain a new matrix $\boldsymbol{\Omega_i}$. We obtain $\boldsymbol{E_R^i} = \boldsymbol{E_i}\boldsymbol{\Omega_i}$ and $\boldsymbol{x_R^i} = \boldsymbol{x_i}\boldsymbol{\Omega_i}$ for zero shrinking and apply SVD as follows:

$$\boldsymbol{E_R^i} = \boldsymbol{U}\boldsymbol{\Delta}\boldsymbol{V}^{\mathrm{T}} \tag{9}$$

After SVD, each atom $\boldsymbol{d_i}$ can be updated by $\boldsymbol{d_i} = \boldsymbol{u_1}$. Then, the sparse coefficient $\boldsymbol{x_R^i}$ is updated by

$$\boldsymbol{x_R^i} = \boldsymbol{\Delta}[1, 1] \cdot \boldsymbol{v_1} \tag{10}$$

Now, the first column of dictionary has been updated, and the above process can be repeated to update the whole dictionary.

III. THE GMM-SR-HSMM FRAMEWORK

A. THE GAUSSIAN MIXTURE MODEL

Cluster analysis [34]–[38] is an approach for exploiting clusters by using the greatest similarity within the same cluster, which is widely used in degradation state recognition. Clustering methods can generally be divided into two categories:

nonparametric methods and probability-based methods. Dean *et al.* [39] compared the two types of methods and found that the probability model-based approaches always have a better clustering performance under the same circumstances.

In this paper, a GMM method combined with an expectation and maximization (EM) algorithm for similarity estimation is introduced. Fig. 2 displays the GMM clustering processes for different distributions. As Fig. 2 shows, the GMM is established by mixing all the data points from different Gaussian distributions together and calculating the probabilities of each data point. Moreover, the GMM parameters are estimated using training data and the iterative EM algorithm. For more details about the GMM, previous paper [40] can be consulted.

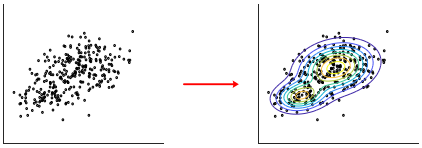


FIGURE 2. The schematic diagram of the GMM model.

B. THE SR FEATURE

A new degradation feature, termed the SR feature, is used in this paper. Although it has been demonstrated to be efficient for fault diagnosis [41], this is the first time that the SR feature has been used as a degradation feature and combined with the HSMM. The process of extracting the SR feature is divided into 3 steps.

(1) The WPT is performed to find the most significant subbands that can characterize the signal information. WPT is a widely used scaled filter that has access to low-frequency information. The low-frequency information has great benefits in the recognition of degradation states. Generally, a good balance of time-frequency resolution is achieved when the signal is decomposed into 4 layers. Thus, 16 subbands $\{\boldsymbol{wp}_1, \boldsymbol{wp}_2, \ldots, \boldsymbol{wp}_{16}\}$ are obtained after the WPT, and we can find the most significant subbands according to the energy proportion of each subband.

(2) For each selected subband, a dictionary $\boldsymbol{D_i}$ can be trained that reflects the fault information well. The K-SVD algorithm is used for training dictionaries and the normal state signal is taken as the set of training samples. In this way, a dictionary set $\{\boldsymbol{D_1}, \boldsymbol{D_2}, \ldots, \boldsymbol{D_\epsilon}\}$ (ϵ denotes the number of subbands selected) can be obtained in the pretraining phase. If we reconstruct the selected subbands of the testing samples with the same dictionary set, different residuals will be obtained.

(3) In this step, we build a signal reconstruction model with the pretrained dictionary set. First, the testing sample is decomposed by WPT. For each subband $\boldsymbol{wp_i}$, we assume the measurement vector is represented as

$$\boldsymbol{\lambda} = \boldsymbol{\Phi} \cdot \boldsymbol{wp_i} \tag{11}$$

In (11), $\boldsymbol{\Phi}$ denotes the measurement matrix. Then, a k-means singular value decomposition (K-SVD) algorithm is applied to $\boldsymbol{wp_i}$. If the testing sample corresponds to the dictionary set, the sparse decomposition will be efficient, and this method will lead to good reconstruction performance. The CS model is established as follows:

$$\boldsymbol{\lambda_i} = \boldsymbol{\Phi_i} \cdot \boldsymbol{wp_i} = \boldsymbol{\Phi_i} \cdot \boldsymbol{D_i} \cdot \boldsymbol{\theta_i} \quad (i = 1, \ldots, \epsilon) \tag{12}$$

In (12), $\boldsymbol{\theta_i}$ denotes the sparse coefficients. The Bayesian compressive sensing (BCS) algorithm has been applied here for the multitask CS reconstruction of $\boldsymbol{\theta_i}$. For an unknown sample, the residuals δ_i of the selected subbands using dictionary $\boldsymbol{D_i}$ can be obtained by

$$\delta_i = \left\| \boldsymbol{wp_i} - \boldsymbol{D_i}\boldsymbol{\theta_i'} \right\|_2 \quad (i = 1, \ldots, \epsilon) \tag{13}$$

Then, the reconstruction residual vector $(\delta_1, \delta_2, \ldots, \delta_\epsilon)$ including the equipment information can be further utilized.

C. THE GMM-SR-HSMM APPROACH

As shown in Fig. 3, the newly developed GMM-SR-HSMM approach contains three main steps.

1) GMM CLUSTERING

GMM clustering is applied to determine the degradation states. Determining the degradation states is the key to gearbox recognition and fault prediction. If the selected number of states is too large, it is difficult to distinguish the various states according to their probability values. If the selected number of states is too small, the results will not be good for judging the fault severity. Generally, we can determine the number of degradation states in two ways. One way is based on experience. For example, Dong and He [42] divided pump degradation into four states according to the amount of oil pollution. The other way is based on indicators. Four typical indicators can be used: the Partition Coefficient (PC) [43], the Partition Index (SC) [44], the Separation Index (SI) [44], and Xie and Beni's Index (XB) [45]. Since the equipment is increasingly complex, the method based on experience lacks versatility and requires a large number of experiments. Therefore, this paper chooses the indicator-based method. Then, the GMM is performed according to the number of degradation states a.

2) SR FEATURE EXTRACTION

After the degradation states are determined, the SR feature needs to be extracted to prepare the HSMM. The feature selection is an important step that directly influences the recognition results. A great many new features have been proposed to improve the degradation assessment. We compare these features with the SR feature in the following.

3) DEGRADATION STATE RECOGNITION AND REMAINING LIFE PREDICTION BASED ON THE HSMM

Degradation state recognition has two parts: HSMM training and HSMM recognition. First, we train the HSMM classifiers

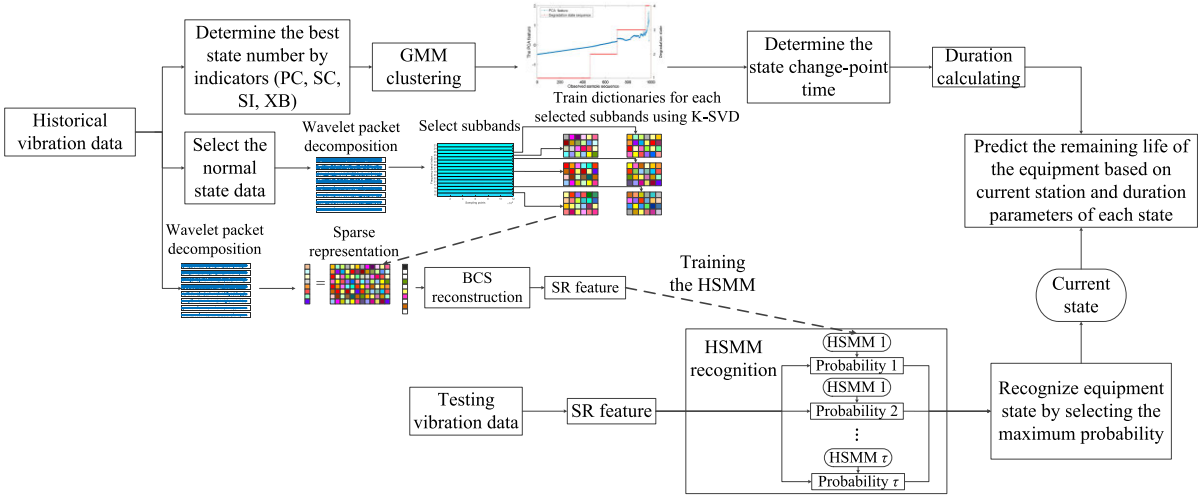


FIGURE 3. The framework of the proposed GMM-SR-HSMM approach.

to recognize the equipment states. In this part, the task is to establish an HSMM for each possible degradation state except for the normal state. In this way, several HSMMs are modeled to characterize each degradation state. Next, we perform HSMM recognition according to the given observation sequence. Each HSMM is applied to the same observation sequence. Then, we can determine the degradation states by choosing the maximum log-likelihood value.

For equipment prognostics, the macrostate duration density $P(x_l)$ can be obtained after HSMM training. Since the density $P(x_l)$ is modeled by a single Gaussian distribution [42], we assume that $\mu(x_l)$ and $\sigma^2(x_l)$ represent the duration mean and variance, respectively, which can be calculated if the duration pdf is obtained. Then, the lower bound, mean value, and upper bound of the duration time can be represented as $\mu(x_l) - \sigma(x_l)$, $\mu(x_l)$, and $\mu(x_l) + \sigma(x_l)$, respectively. For convenience in description, the symbol U is selected to represent the above three values.

In [20], the remaining life prediction process has two parts: determining the probability of each degradation state and determining the RUL value for each state. The RUL at time t can be estimated as follows:

$$RL(t) = \sum_{i=1}^{a} P_i \cdot RL_i \tag{14}$$

$$RL_i = \sum_{j=i}^{a} U_j \tag{15}$$

where P_i denotes the probability of equipment state i, RL_i denotes the estimated RUL of state i, and U_i is the unified variable of the three duration values of state i. From (14) and (15), we can see that the RUL for each state is the sum of the subsequent state durations. Since the entire duration of the state is calculated, the RUL obtained is larger than the actual value. The model is less accurate if we do not divide the degradation states so that the duration of each state is small. On the other hand, if we ignore the state duration, the RUL obtained is smaller than the actual value. Therefore, we try to improve this model by following the method of [46]. The comparison of the original model and improved model is clearly presented in Fig. 4.

$$RL_i = kU_i + \sum_{j=i+1}^{a} U_j \tag{16}$$

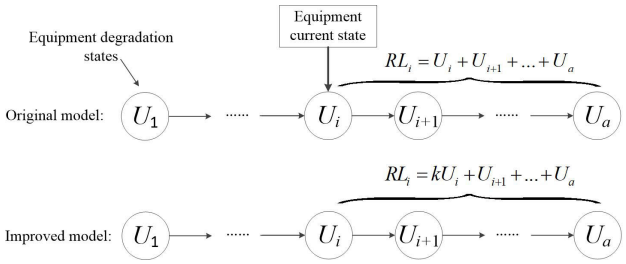


FIGURE 4. The calculation of remaining life for each state.

As shown in (16), we calculate the RUL for each state while considering only a part of the state duration. In this way, the RUL value is more accurate due to introducing the proportion variable k. However, determining the appropriate k for each degradation state is also a challenge. In this paper, an adaptive RUL prediction model is given as follows:

$$RL(t) = \frac{U_{i-1}}{\overline{U_{i-1}}} \sum_{i=1}^{a} P_i \cdot RL_i \tag{17}$$

$$RL_i = k_i U_i + \sum_{j=i+1}^{a} U_j \tag{18}$$

$$k_i = \frac{F_i - F(t)}{F_i - F_{i-1}} \tag{19}$$

In (17), $U_{i-1}/\overline{U_{i-1}}$ is used to adjust the duration of the degradation states. $\overline{U_{i-1}}$ denotes the average duration of the

historical data. It is assumed that different components of the same equipment have a proportional relationship with respect to the degradation duration. In (18), k_i is the variable controlling the duration of the degradation states when calculating an accurate RUL. In (19), F_i denotes the degradation feature of state i, and $F(t)$ represents the degradation feature of time t. The physical meaning of the above variables can be represented as in Fig. 5.

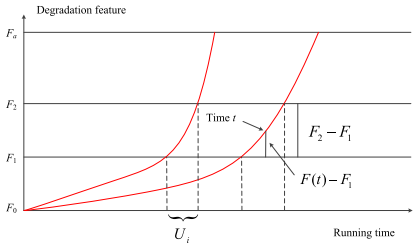


FIGURE 5. Schematic diagram for explaining the prediction parameters.

IV. EXPERIMENTAL VALIDATION

The proposed GMM-SR-HSMM approach is validated using completed-life bearing data [47] from the Center for Intelligent Maintenance Systems (IMS). Fig. 6 shows the bearing test rig. There are four ZA-2115 bearings installed on the shaft. The sampling frequency of the test rig is 20 kHz, and the experiments are conducted at a 2000 RPM rotation speed using an AC motor that is coupled to the shaft via rub belts. An approximately 26671 N radial load is applied to the bearings. The data are collected every 10 minutes and 3 test-to-failure experiments are conducted. We calculated the SR features for each time point and determined the degradation trends for the 3 data sets. Since the degradation trend of the third data set is not obvious, it is not appropriate for RUL prediction. Thus, we use the first two experiments to validate the methods, and they contain 8 test-to-failure data. We denote these 8 data as No. 1 to No. 8.

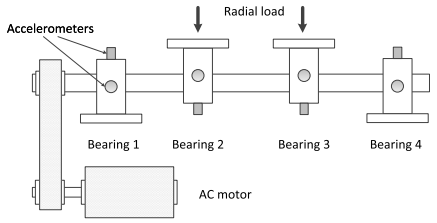


FIGURE 6. Bearing test rig.

At the end of the first experiment, it was found that an inner race defect occurred in bearing 3 and a roller element defect occurred in bearing 4. In all, 2156 files of data were collected in this experiment. At the end of the second experiment, it was found that an outer race failure occurred in bearing 1. In all, 984 files of data were collected in this experiment. Next, the No. 5 bearing is chosen as an example to illustrate our proposed method.

A. STATE DIVISION AND THE DETERMINATION OF THE NUMBER OF STATES

Taking bearing No. 5 as example, we randomly select a number of degradation states c and calculate the values of four indicators: PC, SC, SI, and XB. The formula is introduced in [46]. It should be noted that the larger PC is, the better the performance; and the smaller SC, SI, and XB are, the better the performance. Based on previous experience, the number of degradation states that is set in this paper ranges from 2 to 7. According to Fig. 7, we can easily obtain 4 as the best number of degradation states.

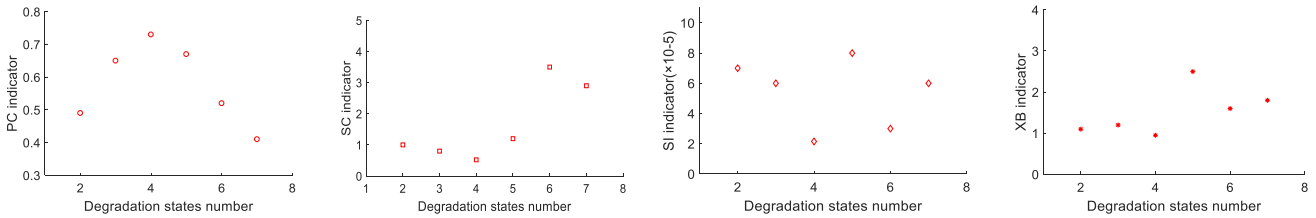
Then, we use the SR feature as the indicator and conduct clustering through the GMM method. Since the GMM allows only a few features, we deal with the 8 channels of the SR features using Principal component analysis (PCA). Finally, one feature can be obtained after PCA, and the GMM is performed on this feature. Moreover, a time factor sequence is also introduced to the clustering to keep the time order. The time length of the IMS data is 984, and so the time factor sequence can be represented as {0/984, 1/984, . . . , 983/984}. The clustering result is shown as Fig. 8.

In Fig. 8, there are 4 states in the transition from normal to failure, which represents the normal state, the early degradation state, the moderate degradation state, and the failure state. We can easily obtain the interval of each degradation state: [1,466], [467,701], [702,951], and [952,984], respectively. It can be seen that the curve is relatively stable in the first two states, and the fluctuation in the third state is randomly low and high. Finally, the observed sequence suddenly rises at approximately 952. According to Fig. 8, the early failure point of 701 is found through the SR feature, which is convenient for the remaining life prediction and health management. Moreover, we can find the functional failure point of 952 for the bearing, which is convenient for the timely warning and prevention of serious accidents.

B. DEGRADATION STATE RECOGNITION

For each degradation state, an HSMM is trained using the SR feature that was extracted from the IMS data. These HSMMs can be represented as HSMM1, HSMM2, HSMM3, and HSMM4. In this paper, half of the data are used as training samples and the remaining data are used as testing samples. The initialization and trained parameters (initial distribution $\boldsymbol{\pi}$ and state transition probability distribution $\boldsymbol{A}$) for each state are listed in Table 1.

Then, the trained HSMMs are used for degradation state recognition. We calculate the probabilities for each HSMM using the same data, and chose the largest probability as the recognized degradation state. The probabilities of certain observation sequences under each HSMM can be calculated by the forward–backward algorithm. The probabilities of the four HSMMs are shown as the four curves in Fig. 9. According to Fig. 9, all of state 1 is correctly recognized. The first halves of state 2 and state 3 are also correctly recognized. However, the boundaries between states 2 and 3 and



(a) The PC indicator (b) The SC indicator (c) The SI indicator (d) The XB indicator

FIGURE 7. Values of the four indicators for different numbers of degradation states.

TABLE 1. Initialization and trained parameters of the HSMM.

Initial parameters	$\boldsymbol{\pi}_0 = [1,\ 0]$	$\boldsymbol{A}_0 = \begin{bmatrix} 0.5, & 0.5 \\ 0, & 1 \end{bmatrix}$
State 1	$\boldsymbol{\pi}_1 = [0.6663,\ 0.3337]$	$\boldsymbol{A}_1 = \begin{bmatrix} 0.9957, & 0.0043 \\ 0.5008, & 0.4992 \end{bmatrix}$
State 2	$\boldsymbol{\pi}_2 = [0.6659,\ 0.3341]$	$\boldsymbol{A}_2 = \begin{bmatrix} 0.9916, & 0.0084 \\ 0.5017, & 0.4983 \end{bmatrix}$
State 3	$\boldsymbol{\pi}_3 = [0.6657,\ 0.3343]$	$\boldsymbol{A}_3 = \begin{bmatrix} 0.9920, & 0.0080 \\ 0.5020, & 0.4980 \end{bmatrix}$
State 4	$\boldsymbol{\pi}_4 = [0.3505,\ 0.6495]$	$\boldsymbol{A}_4 = \begin{bmatrix} 0.4685, & 0.5315 \\ 0.0621, & 0.9379 \end{bmatrix}$

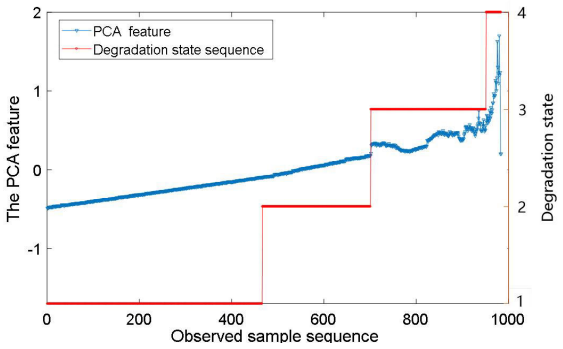


FIGURE 8. Identification of degradation states based on the SR.

states 3 and 4 are recognized with some errors. Moreover, state 4 is nearly correctly recognized. The overall accuracy of the HSMM using the SR feature is calculated as 87.58%.

To demonstrate the efficiency of the SR feature, we compare it with four others: the WPE [14], Renyi entropy [48], NA4 [49] and EEMD [8] features. The WPE is the most commonly used degradation feature in equipment state evaluation and fault detection tasks. The Renyi entropy is a kind of Shannon entropy, which is a feature that measures the amount and complexity of signal information. The complexity of the signal reflects the different states of equipment operation. The NA4 is based on the features proposed by NASA for equipment testing and evaluation. Dempsey and Zakrajsek [49] modified the NA4 to overcome the effect of load changes during the degradation process. Last, EEMD is an effective method for decomposing a time series signal into a series of different frequency bands and is also commonly used.

The definitions of the WPE, kurtosis, Renyi entropy and EEMD features are given in (20), (21), (22) and (23), respectively. In (20), $\boldsymbol{wp}$ denotes the wavelet decomposition subband and *len* denotes the length of $\boldsymbol{wp}$. In (21), parameter α is introduced, and $p_i(x)$ represents the occurrence probability of each value in the discrete sequence. In (22), r denotes the residual signal and H represents the measured signal. Thus, (22) can be regarded as the ratio of the fourth-order statistic of the residual signal to the variance of the current measured residual signal. In (23), $\boldsymbol{\Gamma}_j$ denotes the sample entropy [50] of selected frequency bands, termed the intrinsic mode function (IMF). In addition, compositions of features have been normalized.

$$WPE = \sum_{i=1}^{len} \boldsymbol{wp}(i)^2 \tag{20}$$

$$Renyi = \frac{1}{1-\alpha} In \sum_{i=1}^{N} P_i^{\alpha}(x), \quad \alpha \geq 0,\ \alpha \neq 1 \tag{21}$$

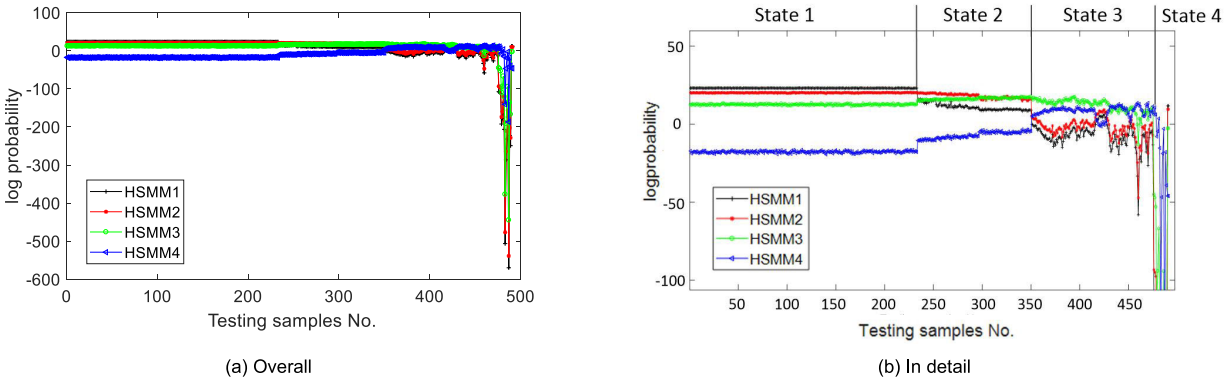
$$NA4 = \frac{G \sum_{i=1}^{G} (r_i - \overline{r})^4}{[(1/H) \sum_{j=1}^{H} \sum_{k=1}^{G} (r_{jk} - \overline{r_j})^2]^2} \tag{22}$$

$$EEMD = (\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \ldots, \boldsymbol{\Gamma}_k)/\boldsymbol{\Gamma}, \quad \boldsymbol{\Gamma} = \sum \boldsymbol{\Gamma}_j \tag{23}$$

Similarly, the HSMMs are established to train the four feature vectors and are used for testing the residual samples. We obtain the accuracy of the HSMM for the WPE, Renyi entropy, NA4 and EEMD features as 54.76%, 75.76%, 41.86% and 71.43% respectively.

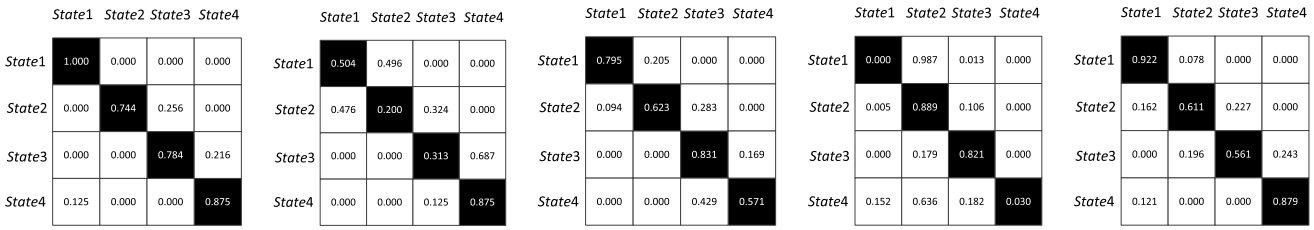
Moreover, the confusion matrices of the HSMM for the five features are given in Fig. 10. Each row of the table represents the distribution of that degradation state. If they are recognized correctly, all samples should be distributed diagonally. Thus, the samples that are not on the diagonal are those with errors.

From Fig. 10, we can see that the recognition results of each degradation state are easily confused with the adjacent states. For example, state 1 is easily confused with state 2, and state 2 is easily confused with states 1 and 3. The recognition results of the HSMM also illustrated this issue. Fig. 9 shows that the error points are concentrated mainly at the state junctions. Therefore, we draw the conclusion that the



(a) Overall (b) In detail

FIGURE 9. Degradation state recognition using the HSMM and the SR feature of the IMS bearing over the full life cycle.



(a) SR (b) NA4 (c) Renyi entropy (d) WPE (e) EEMD

FIGURE 10. Recognition results using the five features.

TABLE 2. The HSMM recognition accuracy based on the five features of the 8 bearings.

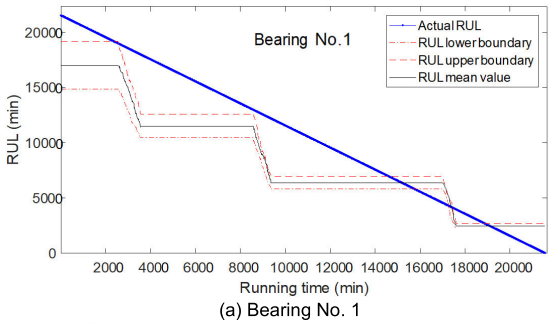
	1	2	3	4	5	6	7	8
SR	80.14%	82.37%	85.64%	83.75%	87.98%	89.92%	88.31%	84.14%
Renyi entropy	68.43%	51.90%	62.39%	61.49%	77.55%	81.34%	80.27%	65.85%
WPE	62.98%	60.33%	67.12%	59.70%	53.87%	75.63%	71.35%	71.28%
NA4	40.56%	40.18%	46.32%	42.21%	40.12%	43.18%	46.23%	44.37%
EEMD	67.20%	53.84%	64.15%	57.76%	70.96%	80.13%	78.52%	69.23%

degradation state transition is the most difficult aspect of recognition.

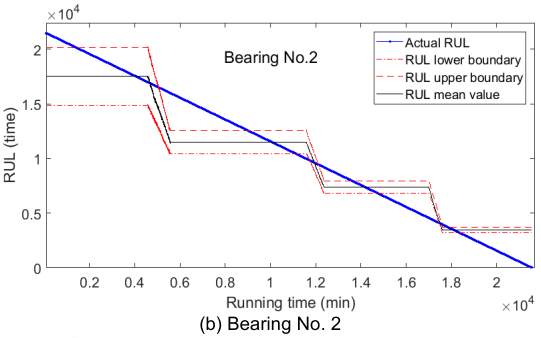
As a further study, 5 groups of HSMMs are trained using the different features of the 8 bearing data from the IMS. As shown in Table 2, in terms of accuracy, the five kinds of features are ranked as follows: SR > Renyi entropy > WPE, EEMD > NA4. There are no obvious conclusions for the order of WPE and EEMD. The accuracy of the SR feature is greater than that of the other four features, achieving an average accuracy of 85.28% for all 8 bearings. The experimental results illustrate that the HSMM based on the SR feature exhibits great advantages in degradation state recognition over the Renyi entropy, WPE, NA4 and EEMD. The classification results confirm that SR features can precisely capture the major structures of the degradation data. The experimental results also show that the NA4 is not suitable for HSMM degradation training in this case because its accuracy is less than 60%. From another perspective, the recognition results of the second group (Nos. 5, 6, 7, and 8) are much better than those of the first group (Nos. 1, 2, 3, and 4). This may be due to the characteristics of the test-to-failure data. The above results indicate that the second set of data is more convenient for degradation state recognition.

C. HEALTH PROGNOSTICS BASED ON HSMM

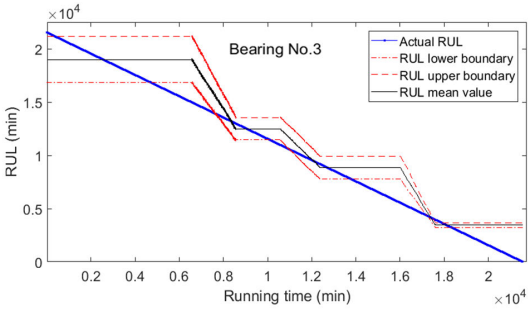
The health prognostics procedure fuses and utilizes the information from the HSMM with the objective of estimating the RUL values. Two health prognostics methods are presented in this section. First, the HSMM is trained offline using the historical lifetime data. After the HSMM training, the transition probabilities between adjacent degradation states can be obtained, as well as the mean and variance of each duration time. Using bearing No. 5 as an example, these values are shown in Table 3. Then, the lower and upper bounds of the predicted state duration can be calculated according to the mean and variance. That is, if the equipment is currently in state i, then its mean RUL, lower-bound RUL and upper-bound RUL can be obtained using the mean, lower bound and upper bound of the predicted state duration, respectively.



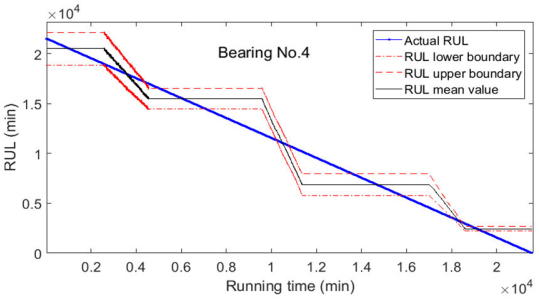
(a) Bearing No. 1



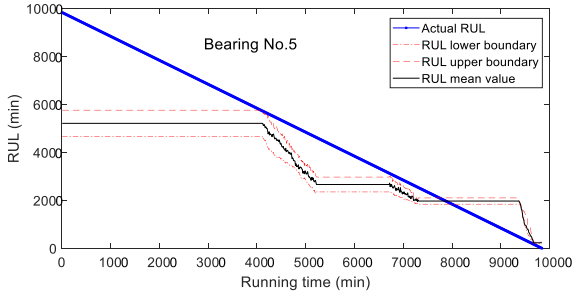
(b) Bearing No. 2



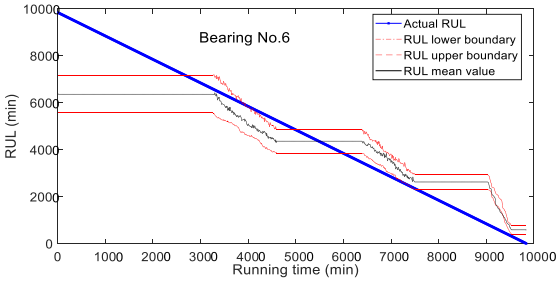
(c) Bearing No. 3



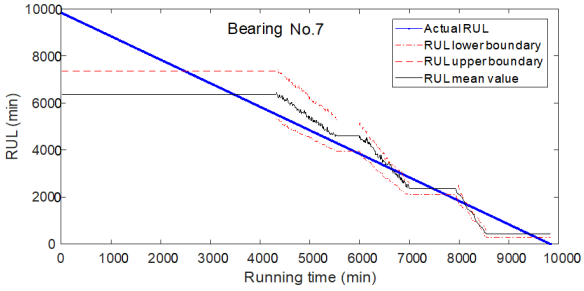
(d) Bearing No. 4



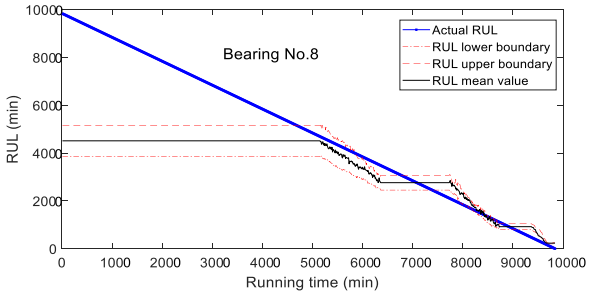
(e) Bearing No. 5



(f) Bearing No. 6



(g) Bearing No. 7



(h) Bearing No. 8

FIGURE 11. RUL prediction of the 8 bearings before revision.

TABLE 3. Mean and variance of the duration for each degradation state.

Degradation state	State 1	State 2	State 3	State 4
Mean	511	260	185	23
Variance	55	31	24	8

The first prediction method can be represented as (14) and (15). Once the degradation state is determined, the RUL of that sample can be calculated based on (14). Fig. 11 shows the RUL prediction using method 1. The blue curve represents the actual RUL, which is a straight line with a slope of 1.

The reason is that the sum of the running time and the RUL at that time is a certain value. As shown in Fig. 11, the RUL line decreases stepwise, and the predicted RUL is quite different from reality. We can also find that the predicted RUL is increasingly more accurate, and the lower bound is closer to the upper bound as the running time increases. To achieve more realistic prediction results, we add some

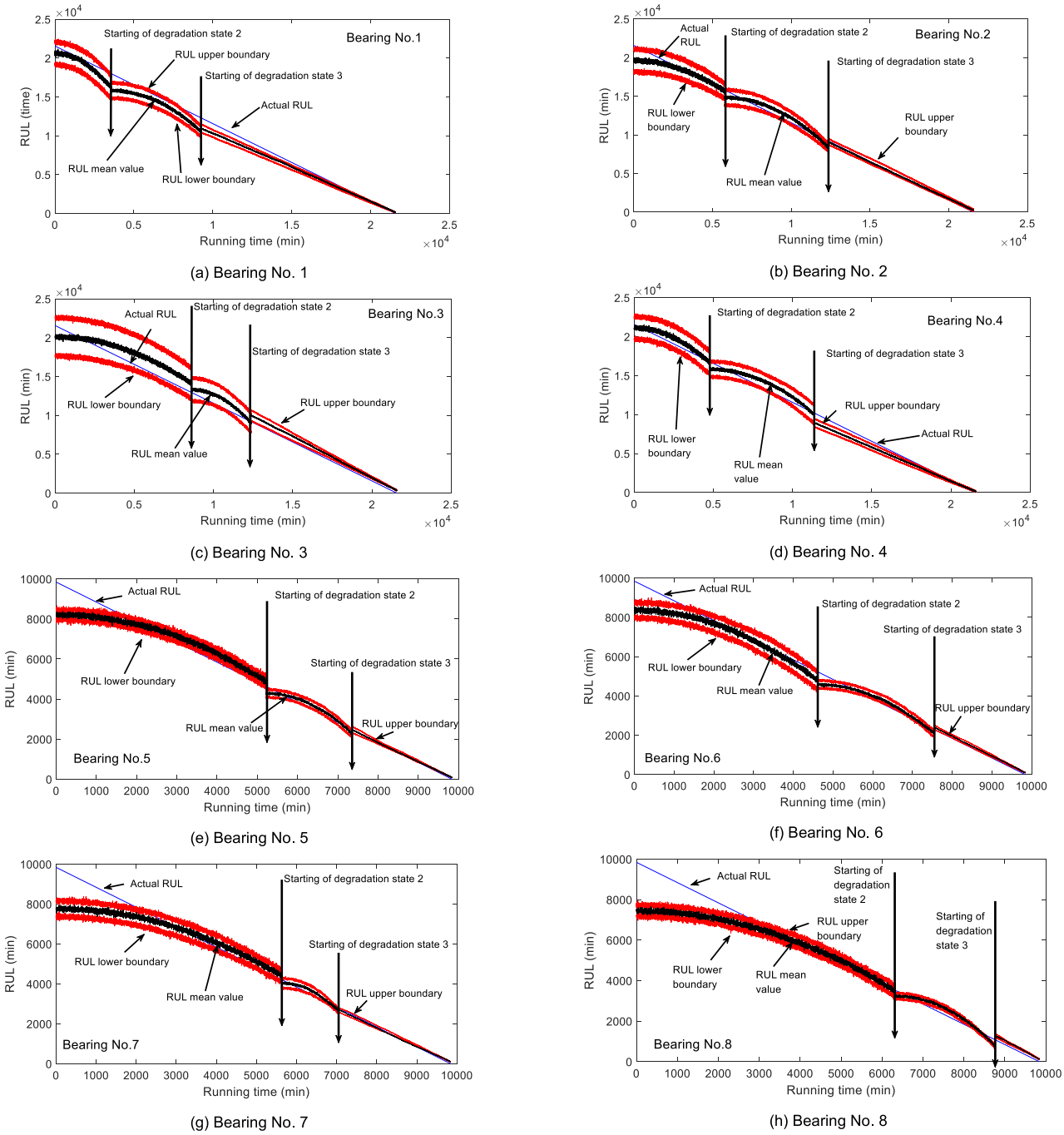


FIGURE 12. **RUL prediction using the method 2 based on the SR feature.**

smoothness between two adjacent states to avoid a sharp change in the predicted curve.

Moreover, this paper presents another prediction method by introducing a control variable k. The value of k can be calculated according to the feature values at the beginning and end of the degradation state. More details can be found in (17)∼(19). To evaluate the performance of the second method, another experiment is conducted. The RUL prediction results based on the improved prediction model of the 8 bearings are shown in Fig. 12. First, we use the proposed SR feature to train the HSMM. It can be seen from Fig. 12 that the curve has a smoother decrease compared to Fig. 11, which indicates that the predicted RUL fits the actual RUL well.

In implementing the improved prediction method, the selection of the duration coefficient k has a great impact on the prediction results. In the 8 groups of bearing data, the values of k can be calculated according to (19). In practice, if k is a certain value, the predicted results will be far

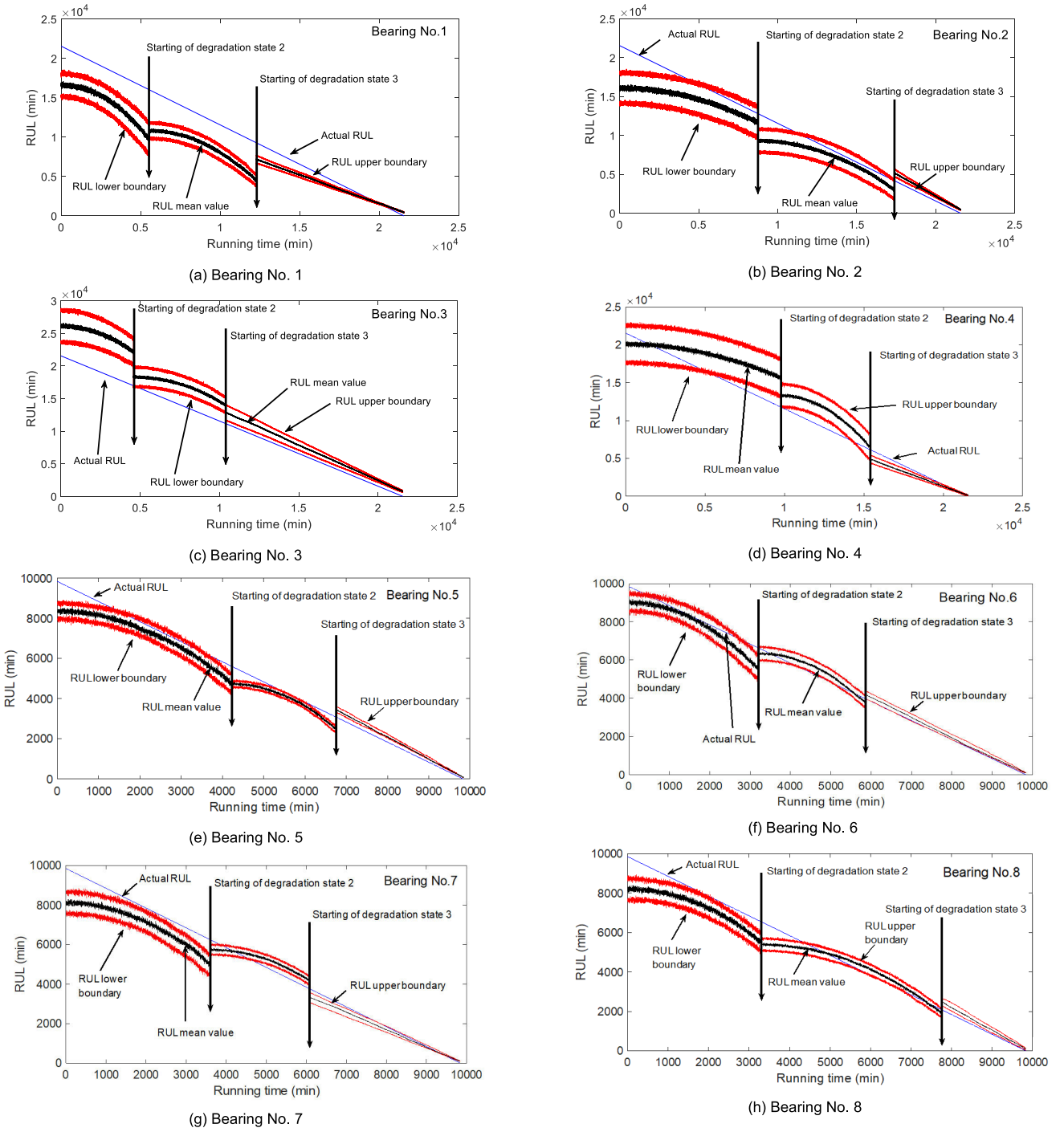


FIGURE 13. RUL prediction using method 2 based on Renyi entropy feature.

from the real values when the actual life of the equipment deviates from the average life. The improved model can adaptively adjust k, and it can be seen from Fig. 12 that the RUL result is more accurate when equipment is in a new degradation state. Moreover, the distance between the lower and upper bounds decreases as the useful life increases, indicating that the prediction error shrinks. This phenomenon has already been confirmed in the experimental results of method 1.

Since the recognition results of Renyi entropy are second only to those of SR, we compare it with SR from a prognostics perspective. The prediction results are shown in Fig. 13. By comparing Fig. 12 and Fig. 13, we find that the SR RUL is closer to the actual RUL value. Therefore, the SR feature

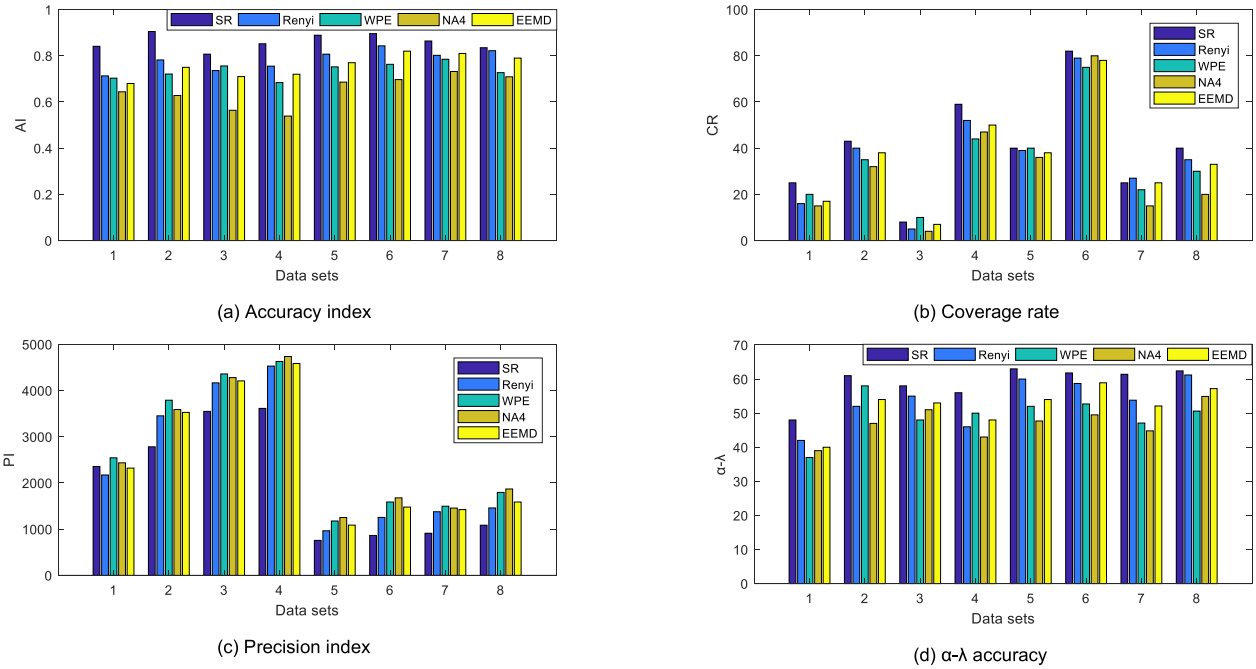


FIGURE 14. **Comparison of the four performance metrics for five kinds of features.**

is more effective for life prediction. Meanwhile, if we compare the two groups of bearing predictions, the results of the second group (that is, bearings 5, 6, 7, and 8) are better. The RUL results for those two features both demonstrate this conclusion. Moreover, the prediction curve changes abruptly when entering degradation states 2 and 3 but not when entering degradation state 1. This is mainly because the prediction results are relatively accurate in state 4, and the change in the formula will not cause a great change in the curve.

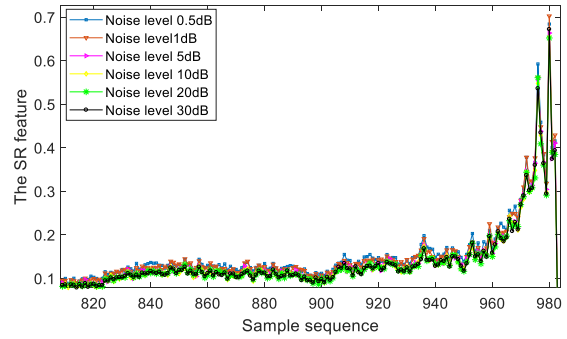
Then, a comparison of the four classical RUL metrics for the Renyi entropy, SR, WPE, NA4 and EEMD is conducted. These four metrics are listed below, and the computation details can be found in Appendix.

a. Accuracy index (AI). This describes the relative error between the actual RUL and predicted RUL values. A smaller AI indicates better prediction performance.
b. Precision index (PI). This describes the width of the estimated RUL value. A smaller PI indicates a better prediction performance.
c. Coverage rate (CR). This describes the probability of the actual RUL within the estimated RUL interval. A larger CR indicates a better prediction performance.
d. α-λ accuracy. This describes the probability of the estimated RUL within a specific confidence interval. A larger α-λ value indicates a better prediction performance. The value of α is set as 10% in this paper.

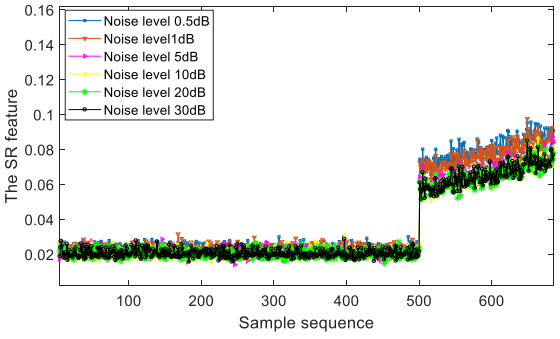
The results in Fig. 14 show that the RUL accuracy indexes of the SR feature are the best compared to those of the other four features. In addition, the average accuracy of SR is 86.11%. The Renyi entropy, WPE, NA4 and EEMD have average accuracies of 78.25%, 73.63%, 64.98%, and 74.12%, respectively. We can see that the ranking of the prediction results follows the ranking of the HSMM recognition. Even though the accuracies are not high for the Renyi entropy, NA4, WPE and EEMD features, the proposed method based on those features was able to predict the remaining life, which benefits PHM development and resilience engineering. The width of the prediction interval provided by the proposed approach is the smallest among the four features, but the coverage interval and α-λ accuracy interval of SR are the highest. The proposed SR feature is more satisfactory in terms of confidence; i.e., it has a larger CR and α-λ accuracy. This is mainly because SR can better extract the structural information of the collected signal and reflect the degradation processes.

Furthermore, the sensitivity of the SR feature is tested. As the HSMM input, the sensitivity of the SR feature can be partly regarded as the sensitivities of the HSMM and the prediction model. From Fig. 15, it can be seen that the change in the SR feature is not obvious as the noise increases. The data intervals from the first stage and the last stage both demonstrate this conclusion.

The case studies show that the proposed SR feature has much better performance than classical degradation features, such as the WPE, Renyi entropy, NA4, and EEMD. The SR feature works well both for degradation recognition and prognostics. From our observations, there are two factors that contribute to the performance improvement in the SR feature. (1) The WPD approach helps the SR to better decompose signals and find the essential structure of the signal, and (2) unlike the classical features, the SR calculates the differences between the degradation signal and the normal signal



(a) The interval [820, 980]



(b) The interval [100, 600]

FIGURE 15. The sensitivity analysis of the SR feature based on different noise levels for different intervals.

so that the feature values of the degradation states are more obvious. Thus, the degradation states are much easier to be distinguished.

V. CONCLUSION

In this paper, we present a GMM-SR-HSMM methodology for equipment health diagnosis and prognosis. (1) The GMM is developed for determining the best degradation states. It should be emphasized that the optimal results obtained by the four indicators may not be completely consistent. (2) The most significant improvement over classical methods is the construction of the degradation feature, which is termed the SR feature. (3) Once the degradation state is determined, the RUL can be easily calculated based on a recursive formula. The choice of the current state residual coefficient k has a great influence on the RUL prediction result. This paper uses an improved model to adaptively adjust the coefficient k.

Compared to the other prediction methods (for example, the neural network method, proportional hazards models, the particle filter model, and MC simulation), the model used in this paper and in [20] is less accurate. However, most prognostic methods still face the problem of providing accurate long-term predictions for industrial applications. Consequently, the approach proposed in this paper can be applied to online machine health prognostics to improve the process reliability during manufacturing processes.

Although the SR feature shows great performance as the input of the HSMM and the prediction model, it requires more time than just calculating the wavelet coefficients. The K-SVD training can be performed offline. However, for achieving CS, the orthogonal matching pursuit (OMP) algorithm is still computationally intense. Optimization methods can be studied to improve the efficiency of the OMP algorithm.

APPENDIX

The performance of RUL prediction can be assessed from different aspects. Four metrics have been adopted in this paper, Accuracy index (AI), Precision index (PI), Coverage Rate (CR), and α-λ accuracy index. Those four metrics can be illustrated in detail as follow:

(1) The Accuracy index (AI) can be calculated as follows:

$$AI = \frac{1}{t}\sum_{t=1}^{T} e^{-\frac{\left|RUL(t)-RUL(t)'\right|}{RUL(t)}} \tag{A.1}$$

where t denotes the whole life time, and $RUL(t)$ and $RUL(t)'$ represent the actual and estimated residual useful lifetimes, respectively.

(2) The Precision index (PI) describes the width of prediction interval, which can be calculated by equation (A.2)

$$PI = \frac{\sum_{t=1}^{T} w_t}{T} \tag{A.2}$$

$$w_t = sup(\widehat{RUL}(t)) - inf(\widehat{RUL}(t)) \tag{A.3}$$

In equation (A.3), $\widehat{RUL}(t)$ represents the set of all estimated residual useful life. Moreover, $sup(\widehat{RUL}(t))$ denotes the upper bound and $inf(\widehat{RUL}(t))$ denotes the lower bound.

(3) The Coverage rate (CR) considers the probability of the prediction interval cover the true value of RUL. It can be represented as

$$CR = \frac{\sum_{t=1}^{T} c_t}{T} \tag{A.4}$$

$$c_t = \begin{cases} 1, & inf\ (\widehat{RUL}(t)) < RUL(t) < sup(\widehat{RUL}(t)) \\ 0, & otherwise \end{cases} \tag{A.5}$$

In equation (A.5), $RUL(t)$ represents the actual and residual useful lifetime. Moreover, $sup(\widehat{RUL}(t))$ denotes the upper bound and $inf(\widehat{RUL}(t))$ denotes the lower bound of the RUL prediction interval.

(4) The α-λ accuracy describes the probability of the predicted RUL fall within specified α-bounds. It can be represented as

$$(\alpha-\lambda) = \frac{\sum_{t=1}^{T} (\alpha-\lambda)_t}{T} \tag{A.6}$$

$$(\alpha-\lambda)_t = \begin{cases} 1, & (1-\alpha)RUL(t) < \widehat{RUL}(t) < (1+\alpha)RUL(t) \\ 0, & otherwise \end{cases} \tag{A.7}$$

In equation (A.7), (α-λ) represents the average α-λ value at the measurement time λ. Larger α-λ value indicates better prediction performance.

REFERENCES

[1] A. Hess and L. Fila, "The joint strike fighter (JSF) PHM concept: Potential impact on aging aircraft problems," in *Proc. IEEE Proc. Aerosp. Conf.*, Big Sky, MT, USA, Mar. 2002, pp. 3021–3026.

[2] N. M. Vichare and M. G. Pecht, "Prognostics and health management of electronics," *IEEE Trans. Compon. Packag. Technol.*, vol. 29, no. 1, pp. 222–229, Mar. 2006.

[3] Z. Liu, Z. Jia, C.-M. Vong, J. Han, C. Yan, and M. Pecht, "A patent analysis of prognostics and health management (PHM) innovations for electrical systems," *IEEE Access*, vol. 6, pp. 18088–18107, 2018.

[4] H. Li and J. P. Ou, "Design and implementation of health monitoring systems for cable-stayed bridges (I): Design methods," *China Civil Eng. J.*, vol. 39, no. 4, pp. 39–44, Apr. 2006.

[5] A. Ljolje and S. E. Levinson, "Development of an acoustic-phonetic hidden Markov model for continuous speech recognition," *IEEE Trans. Signal Process.*, vol. 39, no. 1, pp. 29–39, Jan. 1991.

[6] M. Ostendorf and S. Roukos, "A stochastic segment model for phoneme-based continuous speech recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1857–1869, Dec. 1989.

[7] M.-Y. Chen, A. Kundu, and S. N. Srihari, "Variable duration hidden Markov model and morphological segmentation for handwritten word recognition," *IEEE Trans. Image Process.*, vol. 4, no. 12, pp. 1675–1688, Dec. 1995.

[8] M. Dong, D. He, P. Banerjee, and J. Keller, "Equipment health diagnosis and prognosis using hidden semi-Markov models," *Int. J. Adv. Manuf. Technol.*, vol. 30, nos. 7–8, pp. 738–749, Nov. 2005.

[9] Q. Xiao, Y. Fang, Q. Liu, and S. Zhou, "Online machine health prognostics based on modified duration-dependent hidden semi-Markov model and high-order particle filtering," *Int. J. Adv. Manuf. Technol.*, vol. 94, pp. 1283–1297, Aug. 2017.

[10] F. Cannarile, M. Compare, P. Baraldi, F. D. Maio, and E. Zio, "Homogeneous continuous-time, finite-state hidden semi-Markov modeling for enhancing empirical classification system diagnostics of industrial components," *Machines*, vol. 6, no. 34, p. 34, Jul. 2018.

[11] Q. Liu, M. Dong, and Y. Peng, "A novel method for online health prognosis of equipment based on hidden semi-Markov model using sequential Monte Carlo methods," *Mech. Syst. Signal Process.*, vol. 32, pp. 331–348, Oct. 2012.

[12] Q. Liu, M. Dong, W. Lv, X. Geng, and Y. Li, "A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 217–232, Dec. 2015.

[13] Q. H. Zeng, "Fault prognostics technologies research for key parts and components of mechanical transmission systems," Ph.D. dissertation, Dept. Graduate School, Nat. Univ. Defense Tech., Changsha, China, 2010.

[14] H. Z. Teng, "Research on the state identification of gear box based on CHMM," *J. Vib. Shock*, vol. 31, no. 5, pp. 92–96, May 2012.

[15] X. H. Zhang, "Research of gear box state identification and residual service life prediction based on MoG-HMM," *J. Vib. Shock*, vol. 32, no. 15, pp. 20–25, 2013.

[16] Q. H. Zeng, J. Qiu, and G. J. Liu, "Application of wavelet correlation feature scale entropy and hidden semi-Markov models to equipment degradation state recognition," *Chin. J. Mech. Eng.*, vol. 44, no. 11, pp. 236–247, Nov. 2008.

[17] H. T. Wang, "Prognostic algorithm based on the relative entropy fusion and modified ESN," *J. Vib. Shock*, vol. 38, no. 2, pp. 226–233, 2019.

[18] H. Li, Y. Wang, and B. Wang, "The method of Grey Markov remaining service life predictionspecific to generalized mathematical morphological particle," *J. Vib. Eng.*, vol. 28, no. 2, pp. 316–323, Apr. 2015.

[19] S. Al-Dahidi, F. Di Maio, P. Baraldi, and E. Zio, "Remaining useful life estimation in heterogeneous fleets working under variable operating conditions," *Rel. Eng. Syst. Saf.*, vol. 156, pp. 109–124, Dec. 2016.

[20] H.-E. Kim, A. C. C. Tan, J. Mathew, and B.-K. Choi, "Bearing fault prognosis based on health state probability estimation," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5200–5213, Apr. 2012.

[21] X. C. Dang, P. X. Mao, and Z. J. Hao, "Network traffic clustering algorithm based on quick solution of GMM," *Comput. Eng. Appl.*, vol. 51, no. 8, pp. 96–101, Aug. 2015.

[22] K. Mannepalli, P. N. Sastry, and M. Suman, "MFCC-GMM based accent recognition system for Telugu speech signals," *Int. J. Speech Technol.*, vol. 19, no. 1, pp. 87–93, Nov. 2015.

[23] Y. Pan, J. Chen, and X. Li, "Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy c-means," *Mech. Syst. Signal Process.*, vol. 24, no. 2, pp. 559–566, Feb. 2010.

[24] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[25] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[26] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, vol. 3, 2006, pp. 1433–1452.

[27] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[28] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[29] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[30] X. P. Zhang, N. Q. Hu, L. Hu, L. Chen, and Z. Cheng, "A bearing fault diagnosis method based on the low-dimensional compressed vibration signal," *Adv. Mech. Eng.*, vol. 7, no. 7, Jul. 2015, Art. no. 1687814015593442.

[31] M. Russell and R. Moore, "Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Tampa, FL, USA, Apr. 1985, pp. 5–8.

[32] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Mar. 1989.

[33] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[34] M.-S. Yang, C.-Y. Lai, and C.-Y. Lin, "A robust EM clustering algorithm for Gaussian mixture models," *Pattern Recognit.*, vol. 45, no. 11, pp. 3950–3961, Nov. 2012.

[35] B. Wang, H. R. Li, Q. H. Chen, and B. H. Xu, "Rolling bearing performance degradative state recognition based on mathematical morphological fractal dimension and fuzzy center means," *Acta Armamentarii*, vol. 36, no. 10, pp. 1982–1990, Oct. 2015.

[36] B. Wang, W. Wang, X. Hu, and D. J. Sun, "Degradation condition recognition method based on Gath-Geva fuzzy clustering," *Chin. J. Sci. Instrum.*, vol. 39, no. 3, pp. 21–28, Mar. 2018.

[37] D. Pollard, "Quantization and the method of *k*-means," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 199–205, Mar. 1982.

[38] K.-L. Wu and M.-S. Yang, "Alternative c-means clustering algorithms," *Pattern Recognit.*, vol. 35, no. 10, pp. 2267–2278, Oct. 2002.

[39] N. Dean, T. B. Murphy, and G. Downey, "Using unlabelled data to update classification rules with applications in food authenticity studies," *J. Roy. Stat. Soc., C (Appl. Statist.)*, vol. 55, no. 1, pp. 1–14, Jan. 2006.

[40] E. Nowakowska, J. Koronacki, and S. Lipovetsky, "Clusterability assessment for Gaussian mixture models," *Appl. Math. Comput.*, vol. 256, pp. 591–601, Apr. 2015.

[41] C. Wang, M. Gan, and C. Zhu, "Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory," *J. Intell. Manuf.*, vol. 29, no. 4, pp. 937–951, Sep. 2015.

[42] M. Dong and D. He, "A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology," *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2248–2266, Jul. 2007.

[43] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. Boston, MA, USA: Plenum Press, 1981.

[44] A. M. Bensaid, L. O. Hall, J. C. Bezdek, L. P. Clarke, M. L. Silbiger, J. A. Arrington, and R. F. Murtagh, "Validity-guided (re) clustering with applications to image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 112–123, May 1996.

[45] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.

[46] X. H. Zhang, "Equipment degradation state identification and residual life prediction based on MoG-BBN," *J. Vib. Shock*, vol. 33, no. 8, pp. 171–179, 2013.

[47] *NASA Prognostic Data Repository: Bearing Data Set NSF/UCRC Center for Intelligent Maintenance Systems*. [Online]. Available: http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository.

[48] D. Xu and D. Erdogmuns, "Renyi's entropy, divergence and their nonparametric estimators," in *Information Theoretic Learning*. New York, NY, USA: Springer, 2010, pp. 47–102.
[49] P. J. Dempsey and J. J. Zakrajsek, "Minimizing load effects on NA4 gear vibration diagnostic parameter," NASA, Washington, DC, USA, Tech. Rep. NASA/TM-20010210671, 2001.
[50] A. Widodo, M.-C. Shim, W. Caesarendra, and B.-S. Yang, "Intelligent prognostics for battery health monitoring based on sample entropy," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11763–11769, Sep. 2011.

YUN-FEI MA received the M.Sc. degree from Army Engineering University, Shijiazhuang, China, in 2017, where he is currently pursuing the Ph.D. degree. His current research interests include prognostic and health management (PHM), compressive sensing (CS), and deep learning.

XISHENG JIA received the B.S. and M.S. degrees from the Ordnance Engineering College, China, and the Ph.D. degree from the University of Salford, U.K., in 2001. Since 2001, he has been a Professor with the Ordnance Engineering College. He is currently a Professor and a Ph.D. Supervisor with Army Engineering University. He has published four SCI papers. His main research interests include reliability centered maintenance (RCM) and prognostic and health management (PHM).

QIWEI HU received the Ph.D. degree from the Ordnance Engineering College, Shijiazhuang, China, in 2015. He is currently an Associate Professor with Army Engineering University. His main research interests include maintenance resource integration and optimization and neural networks (NNs).

HUAJUN BAI received the M.Sc. degree from Beijing Jiaotong University, Beijing, China, in 2001. He is currently pursuing the Ph.D. degree with Army Engineering University. His current research interests include prognostic and health management (PHM) and wireless sensor networks (WSNs).

CHIMING GUO received the Ph.D. degree from the National University of Defense Technology, Changsha, China, in 2013. He is currently a Lecturer with Army Engineering University. His main research interest includes maintenance resource integration and optimization.

SHUANGCHUAN WANG received the M.Sc. degree from Air Force Logistics University, Xuzhou, China, in 2016. He is currently pursuing the Ph.D. degree with Army Engineering University. His current research interest includes equipment maintenance effectiveness evaluation.