

Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Fault Diagnosis and Prognosis of Bearing Based on Hidden Markov Model with Multi-Features

Weiguo Zhao, Tiancong Shi, Liying Wang[†]

School of Water Conservancy and Hydropower, Hebei University of Engineering, Handan 056021, China

Submission Info

Communicated by Juan Luis García Guirao

Received December 5th 2019

Accepted January 14th 2020

Available online March 30th 2020

Abstract

A new approach to achieve fault diagnosis and prognosis of bearing based on hidden Markov model (HMM) with multi-features is proposed. Firstly, the time domain, frequency domain, and wavelet packet decomposition are utilized to extract the condition features of bearing vibration signals, and the PCA method is merged into multi-features to reduce their dimensionality. Then the low-dimensional features are processed to obtain the scalar probabilities of each bearing condition, which are multiplied to generate the observed values of HMM. The results reveal that the established approach can well diagnose fault conditions and achieve the remaining life estimation of bearing.

Keywords: hidden Markov model; fault diagnosis; prognosis; multi-features; wavelet packet**AMS 2010 codes:** 68U20; 68T01.**1 Introduction**

As is known to us bearing is one of the most important components used in modern engineering machinery. Once the bearing fails, it will lead to serious consequences such as equipment damage and great economic loss. Fault diagnosis and prognosis for bearing are very important, which can effectively prevent unexpected failures and assist engineering technicians to implement targeted equipment maintenance [1–4]. Fault diagnosis is used for identifying its symptom and fault conditions, and prognosis approach is generally employed to implement the remaining life prediction by existing information and knowledge. Before implementing fault diagnosis and prognosis approaches, it is key for us to effectively extract the fault features of bearing signals, which have direct effects on the diagnosis precision and prediction of bearing. Therefore, the selection for signal features of bearing can comprehensively and concretely reflect the information condition of bearing from different levels [5, 6]. In general, the time-domain component and frequency-domain component can be viewed as important elements

[†]Corresponding author.Email address: wangliying@hebeu.edu.cn

of bearing vibration signals [7,8]. In addition, some low and high frequencies are closely related to the operating states of bearing, which can be obtained by wavelet packet decomposition [9,10]. These components can reflect the detail features of signals from different levels [11]. In this study, the time-domain and frequency-domain components, as well as different frequency scales of vibration signals of bearing are used for the extraction of bearing vibration signals.

After the signal features are extracted, an effective prognostic and prognostic approach needs to be determined. There are many fault diagnosis and prognosis approaches in literature which can be roughly categorized into three categories: experience-based, model-based, and data-driven methods [12]. The experience-based approach only depends on statistical reliability and past experience information, resulting in low accuracy [13]. The model-based approach makes use of complicated physical model and damage propagation model to analyze the degradation process of machines. Though the precision is high, the cost is very high [14]. Data-driven method actually belongs to artificial intelligence approaches, in which the routine condition data of the equipment is used to train a good model. Data-driven methods are viewed as an effective prognostics approach [15]. With the popularity of data-driven approaches, they are increasingly applied to different engineering areas, including artificial neural network (ANN) [16], relevance vector machine (RVM) [17], Bayesian network [18], hidden Markov model (HMM) [19], and support vector regression (SVR) [20]. HMM represents the individual component states of a dynamic system in a natural way. This fact makes this method useful in fault detection and mechanical system monitoring. With its emergence, it has been widely applied as a data-driven modeling approach to different fields such as pattern recognition [21], image identification [22], speech recognition [23] and so on. The intent of this work needs to not only precisely describe the fault conditions of bearing, but also reflect the transformation relation among different fault conditions. Such a dual stochastic process is consistent with the process of HMM. Therefore, the HMM method can be employed to implement fault diagnosis and remaining life prediction of bearing.

In this study, a comprehensive fault diagnosis and prognosis technique based on HMM is proposed. In this approach, the time-domain and frequency-domain, three-layer wavelet decomposition components are extracted from the vibration signals of bearing. Meanwhile, The PCA method is used to fusion multi-features to reduce the dimensionality of the multi-features. Then the scalar probabilities of all the operating conditions of bearing are obtained by the scalar quantization and the observation values of HMM are obtained through the multiplication of these scalar probabilities. According to these observed values, the fault diagnosis and remaining life prediction by HMM can be implemented. The experimental results discover that the proposed scheme not only shows a good fault identification effect for different conditions of bearing, but only makes a remaining life estimation.

2 The Extraction of Multi-features

2.1 Time and Frequency Domains

For the vibration signals of bearing, the time-domain analysis and frequency-domain analysis are two extraction methods of different dimensions. The time-domain analysis represents the dynamic characteristics of vibration signals from the time level. The frequency-domain analysis represents the dynamic characteristics of signals from the frequency level. The time-domain representation, by contrast, is more visual, while the information contained in the signals can be easily observed by frequency-domain representation.

Table 1 provides 16 different time-domain statistical characteristics [24], where $x(n)$ is time-domain signal sequence and n is the number of samples. These time-domain characteristics include the amplitude mean, root mean square, root amplitude, mean root amplitude, degree of skewness, kurtosis, variance, maximum amplitude, minimum amplitude, peak-to-peak value, waveform factor, peak factor, pulse index, margin index, and kurtosis index. Equations ft_1 - ft_{10} reflect the energy and amplitude of time domain signals and equations ft_{11} - ft_{16} reflect the time series distribution of time domain signals.

Table 2 provides 14 frequency-domain statistical characteristics, where $s(k)$ is the spectrum of $x(n)$, k is

Table 1 Time-domain statistical characteristics.

$ft_1 = \frac{1}{N} \sum_{i=1}^N x_i$	$ft_2 = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	$ft_3 = [\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i }]^2$	$ft_4 = \frac{1}{N} \sum_{i=1}^N x_i $
$ft_5 = \frac{1}{N} \sum_{i=1}^N x_i^3$	$ft_6 = \frac{1}{N} \sum_{i=1}^N x_i^4$	$ft_7 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$	$ft_8 = \max\{ x_i \}$
$ft_9 = \min\{x_i\}$	$ft_{10} = \max(x_i) - \min(x_i)$	$ft_{11} = \frac{ft_2}{ft_4}$	$ft_{12} = \frac{ft_8}{ft_2}$
$ft_{13} = \frac{ft_8}{ft_4}$	$ft_{14} = \frac{ft_8}{ft_3}$	$ft_{15} = \frac{ft_5}{ft_3^3}$	$ft_{16} = \frac{ft_6}{ft_2^4}$

Table 2 Frequency-domain statistical characteristics.

$ff_1 = \frac{\sum_{k=1}^K s(k)}{k}$	$ff_2 = \frac{\sum_{k=1}^K (s(k)-ff_1)^2}{k-1}$	$ff_3 = \frac{\sum_{k=1}^K (s(k)-ff_1)^3}{k(\sqrt{p_2})^3}$	$ff_4 = \frac{\sum_{k=1}^K (s(k)-ff_1)^4}{k \cdot ff_2^2}$
$ff_5 = \frac{\sum_{k=1}^K f_k s(k)}{\sum_{k=1}^K s(k)}$	$ff_6 = \sqrt{\frac{\sum_{k=1}^K (f_k - ff_5)^2 s(k)}{k}}$	$ff_7 = \sqrt{\frac{\sum_{k=1}^K f_k^2 s(k)}{\sum_{k=1}^K s(k)}}$	$ff_8 = \sqrt{\frac{\sum_{k=1}^K f_k^4 s(k)}{\sum_{k=1}^K f_k^2 s(k)}}$
$ff_9 = \frac{\sum_{k=1}^K f_k^2 s(k)}{\sqrt{\sum_{k=1}^K s(k) \sum_{k=1}^K f_k^4 s(k)}}$	$ff_{10} = \frac{ff_6}{ff_5}$	$ff_{11} = \frac{\sum_{k=1}^K (f_k - ff_5)^3 s(k)}{kp_6^3}$	$ff_{12} = \frac{\sum_{k=1}^K (f_k - ff_5)^4 s(k)}{kp_6^4}$
$ff_{13} = \frac{\sum_{k=1}^K (f_k - p_5)^{\frac{1}{2}} s(k)}{k\sqrt{p_6}}$	$ff_{14} = \frac{\sum_{k=1}^K (f_k - ff_5)^2 s(k)}{\sum_{k=1}^K s(k)}$		

the number of spectral lines, and f_k is the frequency value of the k th spectral line. These frequency-domain characteristics include the mean of spectrum, root mean square of spectrum, variance of spectrum, skewness index of spectrum, kurtosis index of spectrum, gravity center of spectrum, spectrum dispersion, mean square of spectrum, main frequency variation index, rate of change, skew of frequency, kurtosis of frequency-domain frequency, and ratio of square root, which reflects the time series distribution of time domain signals. Equation ff_1 reflects the magnitude of vibration energy in frequency domain, equations ff_2 - ff_4 , ff_6 - ff_7 and ff_{11} - ff_{14} reflect the dispersion or concentration degree of spectrum, and, equations ff_5 and ff_8 - ff_{10} reflect the position change of main frequency.

2.2 Wavelet Packet Decomposition

Wavelet packet decomposition method is the further development of wavelet decomposition, which provides more abundant signal analysis. Wavelet packet decomposition is able to decompose each detail coefficient vector into two parts, producing a complete binary tree [25, 26]. Wavelet packet is a linear combination of a series of wavelet functions $\phi^i(t)$ which is expressed by

$$\phi_{j,k}^i(t) = 2^{j/2} \phi^i(2^j t - k) \quad i = 1, 2, \dots \tag{1}$$

where i is the frequency factor, j is the scale factor, and k is the translation factor. Any time domain signal can be decomposed by

$$x(t) = \sum_{i=0}^{2i-1} x_j^i(t), \quad i = 1, 2, \dots \tag{2}$$

$$x_j^i(t) = x_{j+1}^{2i-1}(t) + x_{j+1}^{2i}(t) \tag{3}$$

$$x_{j+1}^{2i-1}(t) = Hx_j^i(t) \tag{4}$$

$$x_{j+1}^{2i}(t) = Gx_j^i(t) \quad (5)$$

where $x_j^i(t)$ is the i th frequency band signal at the j th layer wavelet decomposition, $\{h(k), g(k)\}$ are the scale sequences representing the orthogonal low-pass and high-pass filters, by which the decomposition signals in different frequency band ranges are obtained by means of filtering.

When analyzing the vibration signals of bearing using wavelet packet decomposition, it is necessary to determine the type of wavelet package basis function and the number of wavelet decomposition layers. In this work, we adopt 3-layer one-dimensional wavelet packet decomposition based on db1 wavelet and shannon threshold. The energy of wavelet packet reconstruction coefficients on the third layer is extracted as the characteristic value of vibration signals. This decomposition can generate eight wavelet packet reconstruction coefficients. Therefore, combined with 16 time-domain characteristics and 14 frequency-domain characteristics, in total 38 characteristics of vibration signals of bearing in this experiment are used for analysis.

The PCA is used to reduce the dimensionality and extract useful features from 38 statistical characteristics, the obtained low-dimensional features are viewed as the input of the HMM. The procedure of reducing dimensionality is given by

(1) The time-domain and frequency-domain characteristics are obtained by time-domain analysis and frequency domain analysis.

(2) A series of vibration signals of bearing are decomposed using three layer wavelet packet transition, 8 wavelet packet coefficients are obtained as the statistical characteristics of the model.

(3) The PCA is employed to reduce the dimensionality of the statistical characteristics and the new low-dimensional features are obtained as the input of the model, and the mapping matrix of the PCA is restored.

3 Scalar Quantization of Signal Characteristics

Scalar quantization is a technique for machine training and learning to meet certain requirements. The original data is processed according to a certain partition and the new values are assigned, so the continuous values are transformed into discrete values which can be put into the model for training. After the principal characteristics are normalized sequentially, the scalar quantization is carried out. The scalar quantization process used in this work is as follows

$$y = \begin{cases} 1, x < 0 \\ 2, 0 \leq x \leq 0.1 \\ 3, 0.1 \leq x \leq 0.2 \\ \dots \\ 11, 0.9 \leq x \leq 1 \\ 12, x > 1 \end{cases} \quad (6)$$

where x is the original signal and y is the signal characteristic after the scalar quantization.

4 The HMM for Fault Diagnosis and Prognosis

4.1 Hidden Markov Model

The description of the hidden Markov model was created in the 1960s, initially used in speech recognition in the 1970s. By the late 1980s, the hidden Markov model was applied to the analysis of DNA, and then became an important technology in the field of biological information. As people constantly explore and apply this technology, now it has a wide application in many fields such as fault diagnosis, machine learning, automatic driving, natural language processing, and target recognition. The HMM is a statistical model, which can be used to describe Markov process with hidden state. First, the model needs to be trained according to existing

observation values to determine the parameters in the model, and then the observation values are analyzed and identified according to the established model.

Assume $\lambda = (N, M, A, B, \pi)$ is a hidden Markov model(HMM) [27] where, N : number of states in the model.

M : number of different observations for each state.

A : state transition probability matrix $N \times N$.

B : observation probability matrix for each state $N \times M$.

π : initial state probability matrix $1 \times N$.

Markov hypothesis: the future state of a process does not depend on the past state, but only on the present state

$$q_{i+1} = f(q_i) \tag{7}$$

Immobility hypothesis: the state of the system is not related to time

$$P(q_{j+1}|q_j) = P(q_{i+1}|q_i) \tag{8}$$

Output independence hypothesis: the output of the system is only related to the current state of the system

$$P(o_1, o_2, \dots, o_t | q_1, q_2, \dots, q_t) = \prod_{i=1}^t P(o_i | q_i) \tag{9}$$

In this paper, based on the estimation of model parameters and the determination of the location and severity of fault signals by the maximum probability, the forward algorithm, backward algorithm and forward-backward algorithm are adopted. The definition and symbols of each algorithm are as follows.

Forward algorithm: The forward variable is defined as the probability of ending up in any particular state given the first t observations in the sequence

$$\alpha(t, i) = P(o_1, o_2, \dots, o_t, Q_t = q_i | \lambda), 1 \leq t \leq T \tag{10}$$

where T is the length of observations. The forward recursion is expressed as [28]

$$a_j(t + 1) = [\sum_{i=1}^N \alpha_i(t) a_{ij}] B_j(o_{t+1}) \tag{11}$$

Backward algorithm: The backward variable is defined as the probability of observing the remaining observations given any starting point t

$$\beta(t, i) = P(o_{t-1}, o_{t-2}, \dots, o_T, Q_t = q_i | \lambda), 1 \leq t \leq T - 1 \tag{12}$$

The backward recursion is expressed as [28]

$$\beta_j(t) = [\sum_{i=1}^N a_{ij} B_j(o_{t+1})] \beta_j(t + 1) \tag{13}$$

Forward-backward algorithm (Baum-Welch algorithm): Forward-backward algorithm obtains a set of forward probabilities and a set of backward probabilities, which be used to jointly acquire the distribution over states at any specific time t [29]

$$\begin{aligned} \xi_t(i, j) &= P(Q_t = q_i, Q_{t+1} = q_j | O, \lambda) = \frac{\alpha_t(i) A_{ij} B_j(o_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)} \\ &= \frac{\alpha_t(i) A_{ij} B_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) A_{ij} B_j(o_{t+1}) \beta_{t+1}(j)} \quad 1 \leq i, j \leq N, 1 \leq t \leq T \end{aligned} \tag{14}$$

The probability of the model λ and the observation sequence O at time t is defined as

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j), 1 \leq i, j \leq N \quad (15)$$

When training parameters, the initial probability distribution matrix π and state transition probability matrix A are assigned for each state, and these matrixes are updated according to forward-backward algorithm until the requirement of accuracy is met.

$$\pi_i = \gamma_1(i), 1 \leq i \leq N \quad (16)$$

$$A_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, 1 \leq i, j \leq N \quad (17)$$

4.2 The HMM for Fault Diagnosis

The fault diagnosis of bearing refers to identifying the current health status of bearing according to various signals detected in the operation of bearing. A typical diagnostic procedure based on the HMM with multi-features for bearing involves the following steps and is depicted in Fig. 1.

- (1) Obtain the vibration signals of bearing with various states as the training samples.
- (2) Obtain the time-domain characteristics, frequency-domain characteristics, and wavelet packet reconstruction coefficients according to the original vibration signals.
- (3) Reduce the dimensionality of 38 statistical characteristics and obtain the low-dimensional features and the mapping matrix of the training samples.
- (4) Train the HMM classifiers and obtain a set of parameters of each HMM classifier corresponding to a fault state.
- (5) Obtain the vibration signals of bearing with various states as the test samples.
- (6) Obtain the time-domain characteristics, frequency-domain characteristics, and wavelet packet reconstruction coefficients according to the vibration signals to be identified.
- (7) Map 38 statistical characteristics into low-dimensional features according to the mapping matrix obtained from the training process.
- (8) Input the low-dimensional features into each HMM classifier and obtain M probability values.
- (9) Choose the maximum probability of M HMM classifiers and get the corresponding fault states.

4.3 The HMM for Remaining Life Prediction

The whole life data of each degradation mode of bearing is used to train to HMM, which constitutes the life prediction model library. The degradation is non-recoverable and gets worse gradually. The HMM model is viewed as a left-right model, each degradation stage corresponds to a specific state in this HMM model, and the transitions among different states constitute the transition probability matrix. The structure of a M -state left-to-right HMM is described in Fig. 2.

The feature extraction process of bearing vibration signals is performed on all data from the beginning of operation to the current time to obtain a feature vector sequence, and the duration in a state can be expressed as

$$D(S_i) = \sum_{d=1}^{\infty} dp(d_i) \quad (18)$$

Suppose the probability vector of the current state of the system is (p_1, p_2, \dots, p_M) , the remaining life can be given as

$$R = \sum_{j=i}^M \sum_{d=1}^{\infty} dp(d_j) \quad (19)$$

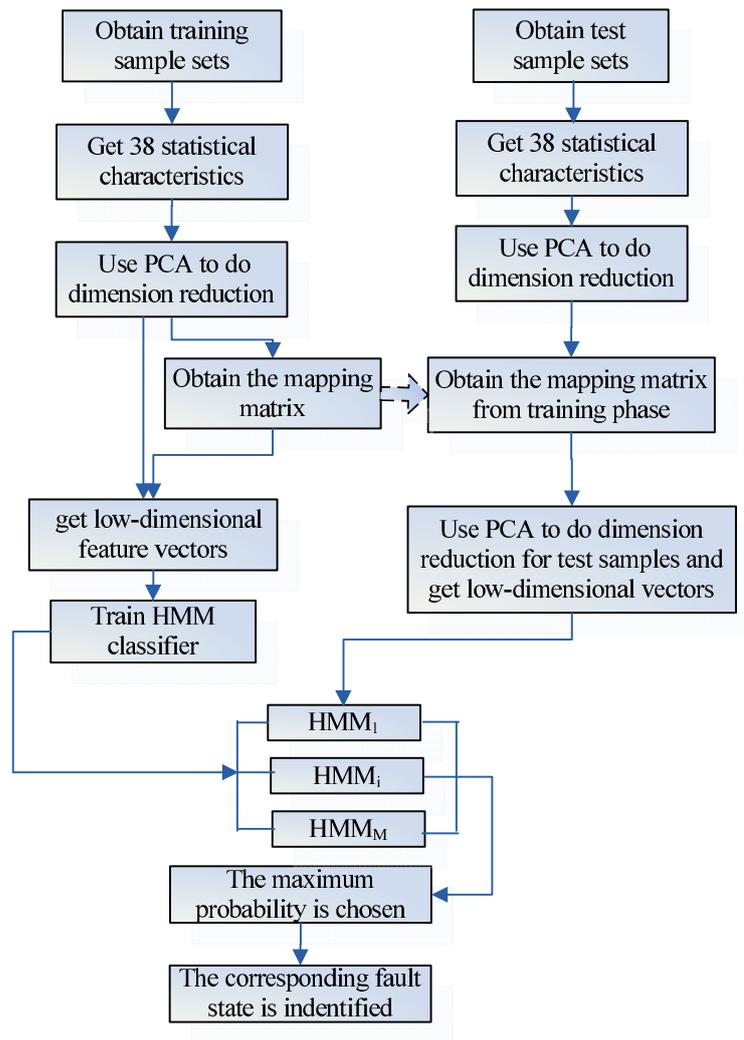


Fig. 1 Proposed scheme for fault diagnosis.

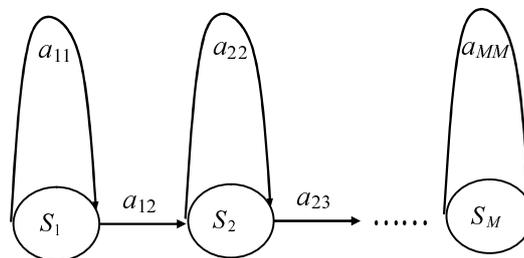


Fig. 2 M-state left-to-right HMM.

Table 3 Some obtained observation values as the input of HMM.

States of faults	No. of observation values									
	1	2	3	4	5	6	7	8	9	10
Nor	0.71	0.03	0.1	0.1	0	0.01	0	0	0.05	0
R1	0.07	0.56	.07	0.1	0	0.17	0	0	0.03	0
R2	0.23	0.08	0.35	0.2	0.01	0.06	0.04	0	0.01	0.02
R3	0.31	0.07	0.06	0.41	0.01	0.03	0.04	0	0.07	0
I1	0	0	0.02	0.01	0.84	0.07	0.01	0	0	0.05
I2	0.04	0.3	0.07	0.06	0.13	0.37	0.02	0.01	0	0
I3	0	0	0.01	0.04	0.11	0.04	0.65	0	0	0.15
O1	0	0	0	0	0	0.02	0	0.97	0	0.01
O2	0.37	0.1	0.02	0.32	0	0	0.01	0	0.18	0
O3	0	0.01	0.03	0.01	0.15	0.03	0.07	0.03	0	0.67

Table 4 The statistics of the overall training results using HMM.

States of faults	Nor	R1	R2	R3	I1	I2	I3	O1	O2	O3
Accuracy	100%	99%	95%	90%	100%	99%	100%	100%	96%	99%

5 Simulation and Analysis

100 samples are obtained for each fault state and the length of each sample is 1024. So there are 1000 training samples corresponding to 10 different fault states are used for training. These training samples are extracted into 38-dimensional characteristic vectors, which are processed by PCA to produce the low-dimensional feature vectors. The dimensionality of feature vectors is determined by

$$\sum_{i=1}^r \lambda_i / \sum_{i=1}^{38} \lambda_i \geq thr, (0 \leq thr \leq 1) \quad (20)$$

where r is the dimension of the feature vectors and thr is a threshold. In this paper the threshold is set as 0.96, and then we can get the projection matrix and $r = 4$.

The 4 characteristic values of each state are changed by scalar quantization according to equation (6), and then these scalar values are multiplied and putted into a probability model, by which the corresponding probabilities from No. 1 to 10 are obtained used for the inputs of the HMM. Table 3 shows some obtained observation values. Nor: normal, R1: slight rotor fault, R2: medium rotor fault, R3: serious rotor fault, I1: slight inter ring fault, I2: medium inter ring fault, I3: serious inter ring fault, O1: slight outer ring fault, O2: medium outer ring fault, O3: serious outer ring fault.

The HMM is established separately for each fault state according the corresponding signals, in which the same probability matrix B is adopted. During the training process, each probability distribution matrix and each state transition probability matrix are redefined according to equations (16) and (17), and the training process is stopped when the convergence precision achieves 0.001. Fig. 3 shows the logarithmic likelihood probability values versus iterations for each fault state. The maximal number of iterations is set to 40. The statistics of the overall training results using HMM are provided in Table 4.

It can be seen that from Fig. 3, the training step for each fault state of HMM is less than 30, demonstrating the strong learning abilities of HMM for the various fault states of bearing. Inspecting Table 4, the training accuracy is relatively high for each fault state, especially for the states of Nor, I1, I3 and O1, their training accuracy comes up to a hundred percent. Fig. 4 shows the fault identification results of training samples.

In order to verify the performance of the established HMM, 250 test samples for each fault state are used for test. After the characteristics of all the test samples are extracted, their corresponding dimensionalities are

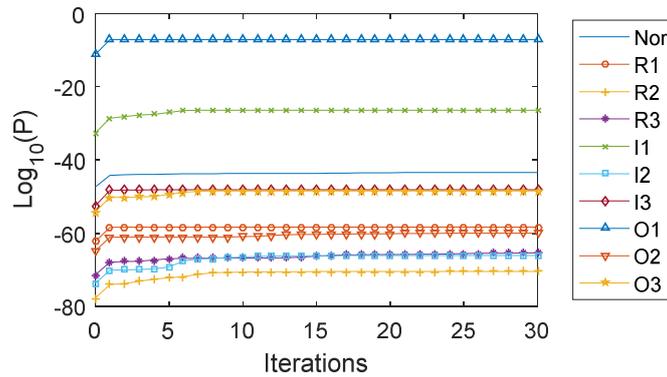


Fig. 3 Training curves of various fault states of HMMs.

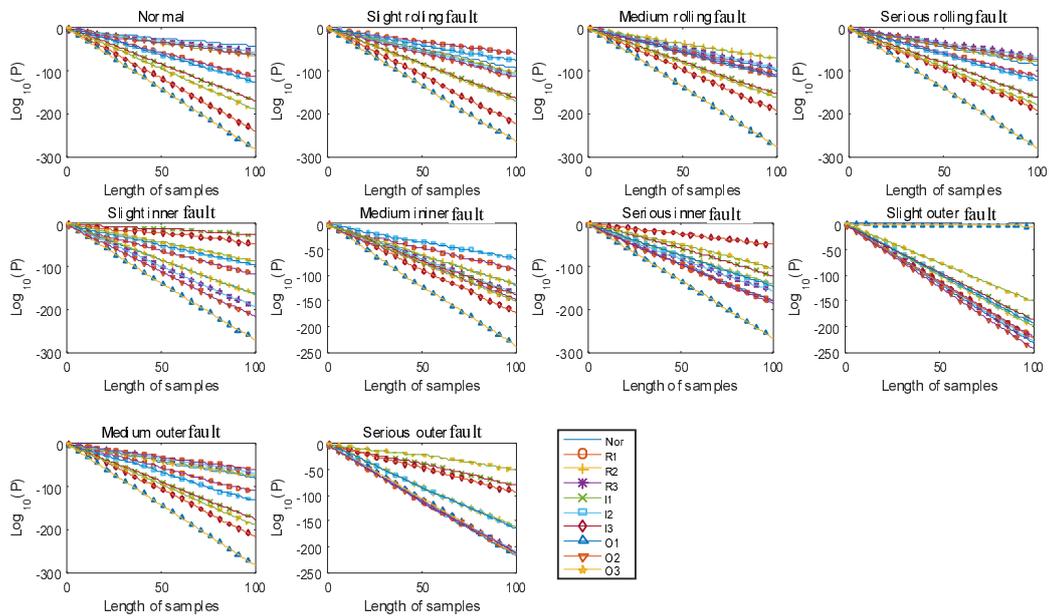


Fig. 4 Fault Identification results of training samples.

reduced and the characteristics are mapped into a low-dimensional space by a mapping vector offered by the training process. Eventually, the processed data is converted into scalar values and putted into the probability model to produce the observation values used for the input of HMM.

To make use of the Markov nature of HMM, the following test processes are implemented.

Process 1: For each fault state, every 10 observation samples are tested as a group, the observation samples from 1 to 10, from 2 to 11, by that analogy, until form 241 to 250, a total of 241 tests can be obtained.

Process 2: For each fault state, every 20 observation samples are tested as a group, the observation samples from 1 to 20, from 2 to 21, by that analogy, until form 231 to 250, a total of 231 tests can be obtained.

The rest can be done in the same manner.

Process 15: For each fault state, every 150 observation samples are tested as a group, the observation samples from 1 to 150, from 2 to 151, by that analogy, until form 101 to 250, a total of 101 tests can be obtained.

According the following processes, the observation sequences are input into 10 HMMs, the likelihood probability values of each model are calculated and the diagnosis result is the state corresponding to the maximum probability. Fig. 5 shows the identification results of various fault states of bearing for every 30 observation

samples using the established HMMs. Fig. 6 shows the identification results of various fault states of bearing for every 100 observation samples using the established HMMs. The statistics of test results of various fault states for different sample lengths are offered in Table 5. As shown in the table, the identification accuracy of various fault states is improved as the length of observation samples increases. The identification result of each fault state reaches the best accuracy when the length of observation samples is more than 120.

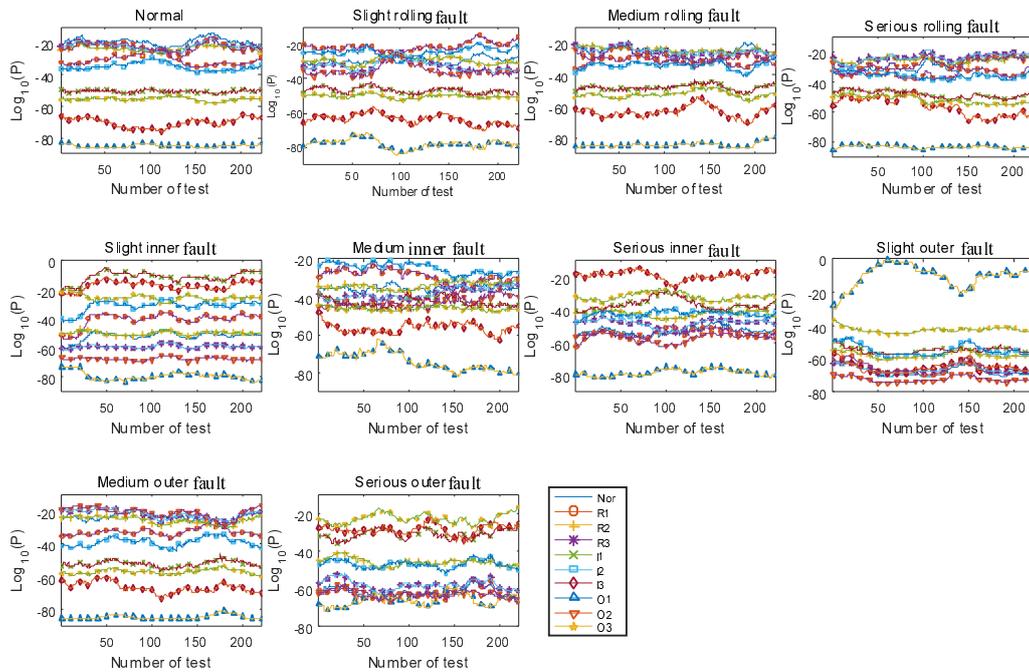


Fig. 5 Identification results of various fault states of bearing for every 30 observation samples.

It is very difficult for us to acquire complete lifetime data of bearing, so the vibration signals are obtained for each degradation phase at intervals. In this simulation, we only investigate the lifetime when the outer race fault appears. We construct 3056 groups of data and the time interval among them is 2 minutes. Each group of data has 1024 sample points. The duration of the normal state is 1450 minutes, the duration of the slight outer race fault is 1368 minutes and the medium outer fault is 238 minutes. Because the performance of the bearings is always getting worse, to reflect the state change of its performance, the HMM is initialized to left-right model with 4 fault states. We assume that the lifetime ends when the serious outer race appears. The half sample data is used for training, the remaining half is used for prediction test. The lifetime curve of prediction test is shown in Fig. 7. From the figure, we can find that the prediction lifetime is improved as the time increases. It is worth mentioning that the prediction results in the practice may be even worse because the actual test data obtained in real situations is imperfect, complex and uncertain.

6 Conclusions

In this study, a diagnostics and prognostics method for bearing, called HMM with multi-domain features, is proposed. First, the multi-features, including time-domain, frequency-domain, wavelet packet decomposition, are used to extract the characteristics from the original vibration signals of bearing. To remove the redundant or irrelevant features, the PCA method is employed for selection features and reduction dimensionality. Then the low-dimensional features are converted into the scalar probabilities multiplied to generate the observation

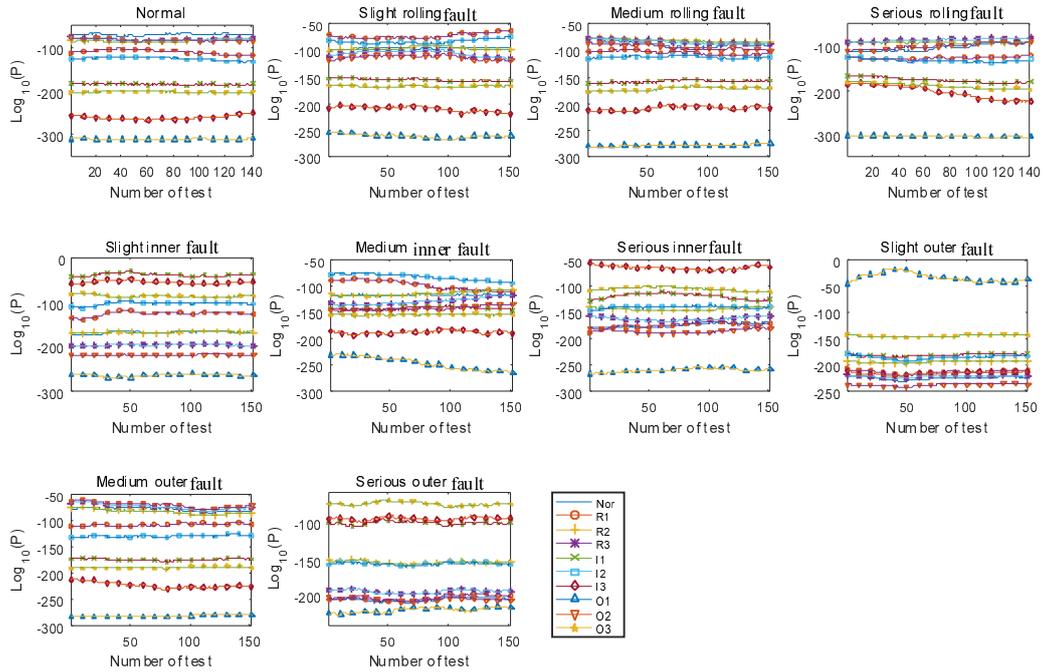


Fig. 6 Identification results of various fault states of bearing for every 100 observation samples.

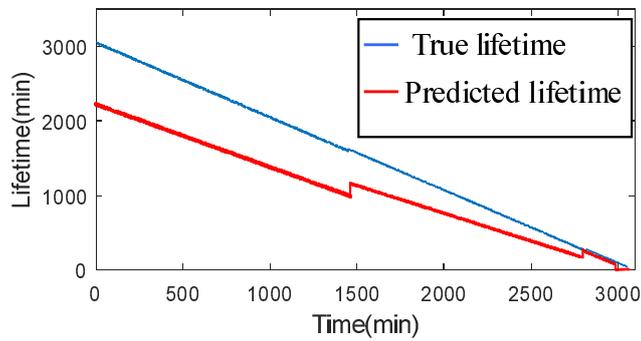


Fig. 7 The lifetime curve of prediction test.

values of HMM, which are fed into HMM model to achieve fault diagnostics and prognostics for bearing. The experimental results show that the proposed scheme is efficient and available in diagnostics and prognostics for bearing. It is very interesting to integrate multi-features with optimized-HMM model for diagnostics and prognostics of bearing [30, 31], this study will be our future focus.

Acknowledgement

This work is supported in part by National Natural Science Foundation of China (Grant: 11972144), Natural Science Foundation of Hebei Province of China (Grant: E2018402092, F2017402142), and Scientific Research Key Project of University of Hebei Province of China (Grant: ZD2017017).

Table 5 Statistics of test results of various fault states for different sample lengths.

Length of observation samples	States of faults									
	Nor	R1	R2	R3	I1	I2	I3	O1	O2	O3
10	64.3%	82.2%	41.1%	53.9%	93.8%	81.3%	97.5%	98.3%	58.9%	85.1%
20	86.1%	88.7%	55.0%	72.3%	93.9%	97.8%	100%	100%	69.7%	92.2%
30	96.8%	95.5%	60.2%	77.8%	99.5%	99.5%	100%	100%	71.9%	98.6%
40	100%	100%	64.0%	77.3%	100%	100%	100%	100%	67.3%	100%
50	100%	100%	63.7%	85.6%	100%	100%	100%	100%	70.6%	100%
60	100%	100%	69.1%	86.9%	100%	100%	100%	100%	74.3%	100%
70	100%	100%	76.8%	90.6%	100%	100%	100%	100%	91.7%	100%
80	100%	100%	83.6%	91.2%	100%	100%	100%	100%	100%	100%
90	100%	100%	83.2%	94.4%	100%	100%	100%	100%	100%	100%
100	100%	100%	90.7%	100%	100%	100%	100%	100%	100%	100%
110	100%	100%	95.7%	100%	100%	100%	100%	100%	100%	100%
120	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
130	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
140	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
150	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

References

- [1] S. Dong and T. Luo, Bearing degradation process prediction based on the PCA and optimized LS-SVM model, *Measurement*, vol. 46, pp. 3143-3152, 2013, DOI: 10.1016/j.measurement.2013.06.038.
- [2] H. Luo, L. Wei, Y. Wang, L. Wang, X. Zhao. A novel approach for analog fault diagnosis based on stochastic signal analysis and improved GHMM. *Measurement*, 81, 26-35, 2016, DOI: 10.1016/j.measurement.2015.11.041.
- [3] S. Benmoussa, M.A. Djeziri, Experimental Application on a Mechanical Transmission System of Integrated Fault Diagnosis and Fault Prognosis method, *IFAC-PapersOnLine*, 51(24), 1016-1023, 2018, DOI: 10.1016/j.ifacol.2018.09.713.
- [4] V.P. Santiago V, Raffaele; Ferrer, Alberto. Data-driven supervised fault diagnosis methods based on latent variable models: a comparative study. *Chemometrics and Intelligent Laboratory Systems*, 187, 41-52, 2019, DOI: 10.1016/j.chemolab.2019.02.006.
- [5] S. Muhammad, K. Cheol-Hong, K. Jong-Myon. A Hybrid Feature Model and Deep-Learning-Based Bearing Fault Diagnosis. *Sensors*, 2017, 17(12):2876, DOI: 10.3390/s17122876.
- [6] Y. Xiaoan, M. Jia. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing* 313(2018) 47-64, DOI: 10.1016/j.neucom.2018.05.002.
- [7] X. Zhang, Z. Liu, J. Wang. Time-frequency analysis for bearing fault diagnosis using multiple Q-factor Gabor wavelets, *ISA Transactions*, 87, 225-234, 2018, DOI: 10.1016/j.isatra.2018.
- [8] H. Huang, N. Baddour, M. Liang. Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve?extraction. *Journal of Sound and Vibration*, 2018, 414:43-60, DOI: 10.1016/j.jsv.2017.11.005.
- [9] G.F. Bin, J.J. Gao, X.J. Li, B.S. Dhillon. Early fault diagnosis of rotating machinery based on wavelet packets-empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing*, 27, 696-711, 2012, DOI: 10.1016/j.ymsp.2011.08.002.
- [10] M.M. Manjural Islam, J.M. Kim. Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network, *Computers in Industry*, 106, 142-1532, 2019, DOI: 10.1016/j.compind.2019.01.00 .
- [11] U.Lepik. Application of Wavelet Transform Techniques to Vibration Studies. *Proc. Estonian Acad. Sci. Phys. Math.*, 2001, 50, 3, 155-168
- [12] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing* 20, 1483-1510, 2006, DOI: 10.1016/j.ymsp.2005.09.012.
- [13] V.T. Tran, B.S. Yang, A.C.C. Tan. Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems, *Expert Systems with Applications* 36 9378-9387, 2009, DOI: 10.1016/j.eswa.2009.01.007.
- [14] M. Roemer, J. Dzakowic, R. Orsagh, C. Byington, G. Vachtsevanos. An Overview of Selected Prognostic Technologies with Reference to an Integrated PHM Architecture. In *Proceedings of the First International Forum on Integrated*

- System Health Engineering and Management in Aerospace, Big Sky, UT, USA, 2005; pp. 3941-3947
- [15] M.A. Djeziri, S. Benmoussa, M.E.H. Benbouzid, Data-driven approach augmented in simulation for robust fault prognosis, *Engineering Applications of Artificial Intelligence*, 86, 154-164, 2019, DOI: 10.1016/j.engappai.2019.09.002.
 - [16] A.K. Jain, J. Mao, K.M. Mohiuddin. Artificial neural networks: A tutorial. *Computer*, 29(3): 31-44, 1996, DOI: 10.1109/2.485891.
 - [17] M.E. Tipping, The relevance vector machine. *Advances in Information Processing System*, 2, 652-658, 2000.
 - [18] N. Friedman, D. Geiger, M. Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3), 131-163, 1997, DOI: 10.1023/A:1007465528199.
 - [19] L.E. Baum, T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *The annals of mathematical statistics*, 37(6), 1554-1563, 1966, DOI: 10.1214/aoms/1177699147.
 - [20] A.J. Smola, B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222, 2004, DOI:10.1023/B:STCO.0000035301.49549.88.
 - [21] T. Chuk, K. Crooke, W.G. Hayward, A.B. Chan, H.J. Hsiao. Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures. *Cognition*, 169, 102-117, 2017, DOI: 10.1016/j.cognition.2017.08.003.
 - [22] K.M. Sagayam , D.J. Hemanth. ABC algorithm based optimization of 1-D hidden Markov model for hand gesture recognition applications. *Computers in Industry*, 99, 313-323, 2018, DOI: 10.1016/j.compind.2018.03.035.
 - [23] J.T. Chien, H.C. Wang, Telephone speech recognition based on Bayesian adaptation of hidden Markov models, *Speech Communication*, 22(4), 369-384, 1997, DOI: 10.1016/S0167-6393(97)00033-2.
 - [24] S. Dong, T. Luo. Bearing degradation process prediction based on the PCA and optimized LS-SVM model. *Measurement*, 46: 3143-3152, 2013, DOI: 10.1016/j.measurement.2013.06.038.
 - [25] N.G. Nikolaou, I.A. Antoniadis. Rolling element bearing fault diagnosis using wavelet packets. *NDT and E International*, 2002, 35(3):197-205, DOI: 10.1016/S0963-8695(01)00044-5.
 - [26] B. Asgarian, V. Aghaeidoost, H.R. Shokrgozar. Damage detection of jacket type offshore platforms using rate of signal energy using wavelet packet transform. *Marine Structures*, 45, 1-21, 2016, DOI: 10.1016/j.marstruc.2015.10.003.
 - [27] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, 1989, 77:257-286.
 - [28] P. Baruah, R.B. Chinnam. HMMs for diagnostics and prognostics in machining processes. *Int. J. Production Res.* 2005, 43, 1275-1293, DOI: 10.1080/00207540412331327727.
 - [29] L.E. Baum, T. Petrie, G. Soules, N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, 41:164-171, 1970, DOI: 10.1214/aoms/1177697196.
 - [30] W. Zhao, L. Wang, An effective bacterial foraging optimizer for global optimization. *Inf. Sci.*, 2016, 329: 719-735.
 - [31] W. Zhao, L. Wang, Z. Zhang, A novel atom search optimization for dispersion coefficient estimation in groundwater, *Future Gener. Comput. Syst.*, 2019, 91: 601-610.

This page is intentionally left blank