Dynamic Diagnosis Approach of Multi-state Degradation System Using Hidden Markov Model

Guangqi Qiu
School of Mechanical and Electrical Engineering
Jiangxi University of Science and Technology
Ganzhou, P.R. China
qiuguangqi@sina.cn

Yingkui Gu
School of Mechanical and Electrical Engineering
Jiangxi University of Science and Technology
Ganzhou, P.R. China
Corresponding author: guyingkui@163.com

Abstract—**A methodology for developing dynamic diagnosis of multi-state degradation system was proposed in this paper. Wavelet packet energy entropy was employed to characterize the uncertainty and complexity of the signal. Current state evaluation and multi-state recognition had been implemented by hidden Markov model. The recognition performance was verified by a bearing vibration experiment, and the effects of decomposition levels and wavelet mother functions on the recognition performance were taken into account. Compared with classifiers of *K*-means, BP neural networks (BP-NN) and support vector machine (SVM), hidden Markov model (HMM) achieved a better recognition performance for multi-state degradation system and provided theoretical explanation of the system failure evolution.**

Keywords- multi-state degradation system; dynamic diagnosis; state recognition; hidden Markov model; wavelet packet energy entropy

I. INTRODUCTION

The system and components are regarded as only working state and failure state in conventional binary reliability theory, ignoring multi-performance system and multiple failure modes. In addition, the performance of the equipment becomes worse gradually with the increasing of run time, and the degradation process of mechanical equipment is discredited as limited discrete-states. Therefore, conventional binary reliability theory is extended to multi-state reliability theory for better describing the deterioration law of mechanical equipment.

The main research objectives of the article are to recognize states of multi-state degradation system by the multi-state information which may reveal the potential failure mechanism, and provide theoretical explanation of the system failure evolution. Unfortunately, two challenges are facing in state recognition of multi-state degradation system: the sensitivity of fault features and the availability of diagnosis model.

To overcome the first challenge, many signal processing technologies, such as fourier transform, time-domain analysis, frequency-domain analysis, wavelet packet energy spectrum [1], independent component analysis [2], principal component analysis [3-4], rough set [5] and mutual information [6], and so on, had been effectively employed to extract sensitive fault features so far. Most notably, Nguyen et al. [7] applied empirical mode decomposition and wavelet packet decomposition method to feature extraction from bearing acoustic emission signal. Rai et al. [8] reviewed three stage of the bearing fault signal processing approaches, and pointed out the recent developments. Although these methods achieved a good performance in many industries, there were still many challenges required more advance techniques, for example, to eliminate the noise, to remove redundant information, and to extract weak features [9].

The another important challenge is how to recognize states of multi-state degradation system and provide theoretical explanation for failure evolution by the appropriate performance assessment model. A variety of artificial intelligence approaches, such as long short-term memory recurrent network [10], random matrix single ring machine learning [11], deep learning [12], and so on, have been applied in bearing performance assessment. Caesarendra et al. [13] combined relevance vector machine and logistic regression method. Zhu et al. [14] combined support vector data description and rough set. Yu [15] combined locality preserving projections and Gaussian mixture models. Li et al. [16] proposed degradation-hidden-Markov model. Pham et al. [17] proposed proportional hazard model and support vector machine for performance assessment and residual life prediction. Compared with the black-box models such as artificial intelligence approaches, HMM is easier to model interpretation [18].

For the above reasons, this paper focuses on the diagnosis approach of multi-state degradation system with wavelet packet energy entropy and hidden Markov model. The accuracy and efficiency of the proposed dynamic diagnosis method were validated by a bearing vibration experiment. The effects of different decomposition levels and wavelet mother functions on the recognition performance were taken into account.

II. STRUCTURE FUNCTION OF MULTI-STATE DEGRADATION SYSTEM

For conventional reliability models, a binary state hypothesis is made in describing mechanical equipment or

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

systems, that is, the system is in either of two possible states: working or failure state. In this condition, the effect of accumulated degradation on system performance is not taken into account, and it is much difficult to produce a enough well result for reliability assessment in actual situation [19]. Generally, the mechanical equipment tends to deteriorate over time owing to stress and load, therefore the performance of the mechanical equipment or the systems is in a gradual degradation process. Figure 1 describes a multi-state degradation system.

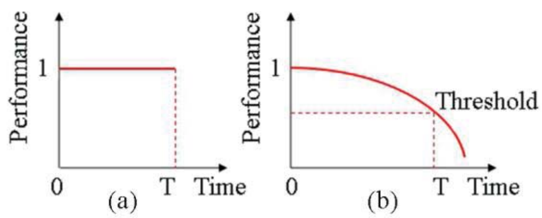


Figure 1. Multi-state degradation system

From Figure 1(a), we can see that the system is considered as two possible states, one is working state and the other is failure state. During the mission time T, the system is in working state, and the system is in failure state after the mission time T. Conventional reliability assessment methods are based on this theory. However, in practice, the most systems or components are exposed to a performance degradation process. As shown in Figure 1(b), the performance of the mechanical equipment is 1 at time $t=0$, and is gradually deteriorated during operational time $0 \le t \le T$. There is a fixed fault threshold at operational time $t=T$. The system fails when it reaches the threshold. In this case, the system is considered as a multi-state degradation system, and its structure function is given in this section.

Let a multi-state degradation system consist of n components: $x_i, (i=1,2,\cdots,n)$. Each component includes m_i+1 states: $(0,1,\cdots,m_i)$, indicating the best state to the worst state from left to right respectively, i.e. $0 \to m_i$, where, 0 denotes the perfect state, and m_i denotes the worst state, the others are degradation states. Structure function $\varphi(x)$ is described as.

$$\begin{aligned}\varphi(x):(0,1,\cdots,m_1)\times(0,1,\cdots,m_2)\times\cdots\times(0,1,\cdots,m_i)\times\cdots \\ \times(0,1,\cdots,m_n)\to(1,2,\cdots,N)\end{aligned} \tag{1}$$

where, the binary system is a special case when the system is only considered as failure state and working state.

III. Dynamic Diagnosis Approach of Multi-State Degradation System

HMM is a statistical method with a set of hidden states, and is actually a double stochastic process. One is a probabilistic state transition process of Markov model, and the another is a stochastic process from the underlying state to the observations. [20]. HMM is applied in speech recognition originally and achieved a good result. HMM is employed for dynamic diagnosis of multi-state degradation system in this paper.

A. *HMM Algorithm*

Figure 2 shows a failure evolution process. To our knowledge, a machine goes through n limited discrete-states before failure, and let state 1 denote the best machine performance, state n denote the worst machine performance. As the state number increases, the machine performance becomes worse. Once the system is described as a HMM, fault modes (or performance) are considered as observation states, operational conditions are considered as hidden states.

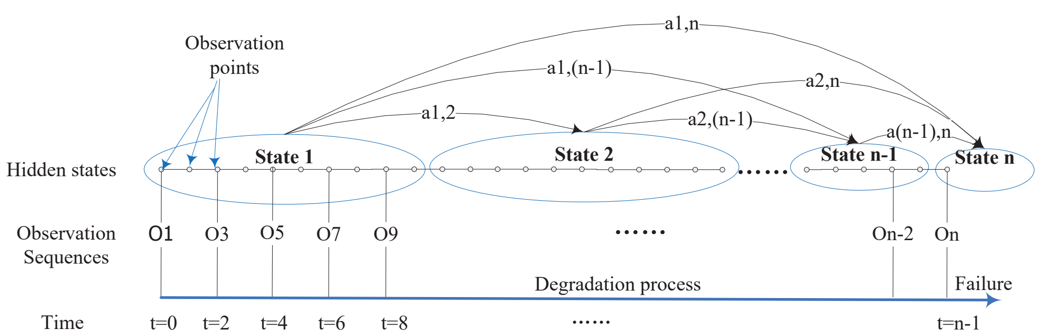


Figure 2. Failure evolution process

Generally, a HMM is defined as $\lambda=\{\pi, A, B\}$ [21].

(1) state number N. The individual states are defined as $\{S_1, S_2, \cdots, S_N\}$.

(2) observation sequence number M. The observation sequence is defined as $\{O_1, O_2, \cdots, O_M\}$.

(3) Initial state distribution $\pi=\{\pi_i, i=1,2,\cdots,N)$, and $\pi_1=1, \pi_i=0 (2 \le i \le N)$.

(4) State transition matrix $A=\{a_{ij}\}$, $A=a_{i,j}=p(S_j \mid S_i), 1 \le i, j \le n$ and $\sum_{j=1}^{n} a_{i,j}=1$. In the real

world, the performance of the equipment becomes worse gradually. It is assumed that $a_{i,j} = 0, \quad if \quad i > j, a_{NN} = 1$.

(5) Confusion matrix $B = \{b_i(k)\}(1 \le i \le M, 1 \le k \le N)$, and $\sum b_i(k) = 1$.

Three important issues of HMM are taken into account [22]:

(1) Evaluation (or Classification): if we know a HMM λ and a observation sequence $O = \{O_1, O_2, \cdots, O_M\}$, how to computer the probability $P(O/\lambda)$.

(2) Decoding (or Recognition): if we know a HMM λ and a observation sequence $O = \{O_1, O_2, \cdots, O_M\}$, how to obtain a hidden state sequence $S = \{S_1, S_2, \cdots, S_N\}$ that most probably generates the given observation sequence.

(3) Learning (or Training): how to obtain the best model $\lambda = \{\pi, A, B\}$ to maximize $P(O/\lambda)$.

The evaluation issue has been solved by forward–backward algorithm.

$$\begin{aligned} P(O \mid \lambda) &= \sum_S P(O \mid S, \lambda) P(S \mid \lambda) \\ &= \sum \pi_1 b_1(O_1) a_{12} b_2(O_2) \cdots a_{(T-1),T} b_T(O_T) \end{aligned} \tag{2}$$

Forward procedure is defined by

$$\alpha_t(i) = P(O_1, O_2, \cdots, O_t, S_t = \theta_i \mid \lambda), 1 \le t \le T \tag{3}$$

where, $\alpha_t(i)$ denotes the probability of the sequence $O_1, O_2, \cdots, O_t$ being in state i at time t. We can recursively define $\alpha_t(i)$ as.

$$\alpha_1(i) = \pi_i b_i(O_{1)} \tag{4}$$

$$\alpha_{i+1}(i) = [\textstyle\sum_{i=1}^N \alpha_t(i) a_{ij}] b_j(O_{t+1}), \begin{matrix} 1 \le t \le T \\ 1 \le j \le N \end{matrix}. \tag{5}$$

$$P(O \mid \lambda) = \textstyle\sum_{i=1}^N \alpha_T(i). \tag{6}$$

Backward procedure is similarly defined by

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \cdots, O_T \mid S_t = \theta_i, \lambda), 1 \le t \le T-1 \tag{7}$$

where, $\beta_t(i)$ is the probability of the sequence $O_{t+1}, O_{t+2}, \cdots, O_T$ being in state i at time t, $\beta_T(i) = 1$. We can efficiently define recursively $\beta_t(i)$ as:

$$\beta_T(i) = 1, 1 \le i \le N \tag{8}$$

$$\beta_t(i) = \textstyle\sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \begin{matrix} t = T-1, T-2, \cdots, 1 \\ 1 \le i \le N \end{matrix} \tag{9}$$

$$P(O \mid \lambda) = \textstyle\sum_{i=1}^N \beta_t(i) \tag{10}$$

The decoding issue has been solved by Viterbi algorithm.

$$\delta_t(i) = \max_{S_1, S_2, \cdots, S_{t-1}} P(S_1, S_2, \cdots, S_{t-1}, S_t = \theta_i, O_1, \cdots, O_t \mid \lambda) \tag{11}$$

where, $\delta_t(i)$ is the probability of the state sequence O being in state i at time t. We can efficiently define recursively $\delta_t(i)$ as.

$$\begin{cases} \delta_1(i) = \pi_i b_i(O_1), 1 \le i \le N \\ \varphi_1(i) = 0, 1 \le i \le N \end{cases} \tag{12}$$

$$\begin{cases} \delta_t(i) = \max\limits_{1 \le i \le N}([\delta_{t-1}(i) a_{ij}]) b_j(O_t), 2 \le t \le T, 1 \le j \le N \\ \varphi_t(j) = \arg\max\limits_{1 \le i \le N}[\delta_{t-1}(i) a_{ij}], 1 \le t \le T, 1 \le j \le N \end{cases} \tag{13}$$

$$\begin{cases} P^* = \max\limits_{1 \le i \le N}[\delta_T(i)] \\ S_T^* = \arg\max\limits_{1 \le i \le N}[\delta_T(i)] \end{cases} \tag{14}$$

The optimal state sequence is

$$S_t^* = \varphi_{t+1}(S_{t+1}^*), t = T-1, T-2, \cdots, 1 \tag{15}$$

The learning issue has been solved by Baum–Welch algorithm

$$\xi_t(i, j) = P(O, S_t = \theta_i, S_{t+1} = \theta_j \mid \lambda) \tag{16}$$

where, $\xi_t(i, j)$ is the probability of the state sequence O being in state i at time t and state j at time $t+1$.

$$\xi_t(i, j) = [\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)] / P(O \mid \lambda) \tag{17}$$

$$\xi_t(i) = P(O, S_t = \theta_i \mid \lambda) = \textstyle\sum_{j=1}^N \xi_t(i, j) = [\alpha_t(i) \beta_t(i)] / P(O \mid \lambda) \tag{18}$$

By forward–backward, Viterbi, and Baum–Welch algorithms, we obtain the following re-evaluation formulas for initial state distribution, state transition and confusion matrices.

$$\overline{\pi}_i = \xi_1(i) \tag{19}$$

$$\overline{a}_{ij} = \sum_{i=1}^{T-1} \xi_t(i,j) / \sum_{i=1}^{T-1} \xi_t(i) \tag{20}$$

$$\overline{b}_i(k) = \sum_{i=1,O_t=o_k}^{T-1} \xi_t(i,j) / \sum_{i=1}^{T-1} \xi_t(j), and O_t = o_k \tag{21}$$

B. Dynamic Diagnosis Procedure of Multi-State Degradation System

Figure 3 shows a dynamic diagnosis method for multi-state degradation system. There are several modules: data processing, feature extraction with wavelet packet energy entropy, scalar quantization, HMM training and state recognition.

(1) Data processing

Restricted to inherent characteristics of the sensors, such as limited precision and limited range, the raw signal must be processed by eliminating the outliner and zero-mean normalization before feature extraction.

(2) feature extraction with wavelet packet energy entropy

Wavelet packet decomposition has been proven to be a high time-frequency resolution signal processing method. Meanwhile the uncertainty and the complexity of the signals can be effectively characterized by information entropy. For wavelet packet decomposition method, the signal is decomposed into the low-frequency and high-frequency component with low-pass and high-pass filters, respectively. The signal $x(t)$ has 2^j sub-band by the j_{th} level of decomposition. The sub-band width Δf, sample frequency f_s and decomposition level j satisfy the following relation.

$$\Delta f = f_s / 2^{j+1} \tag{22}$$

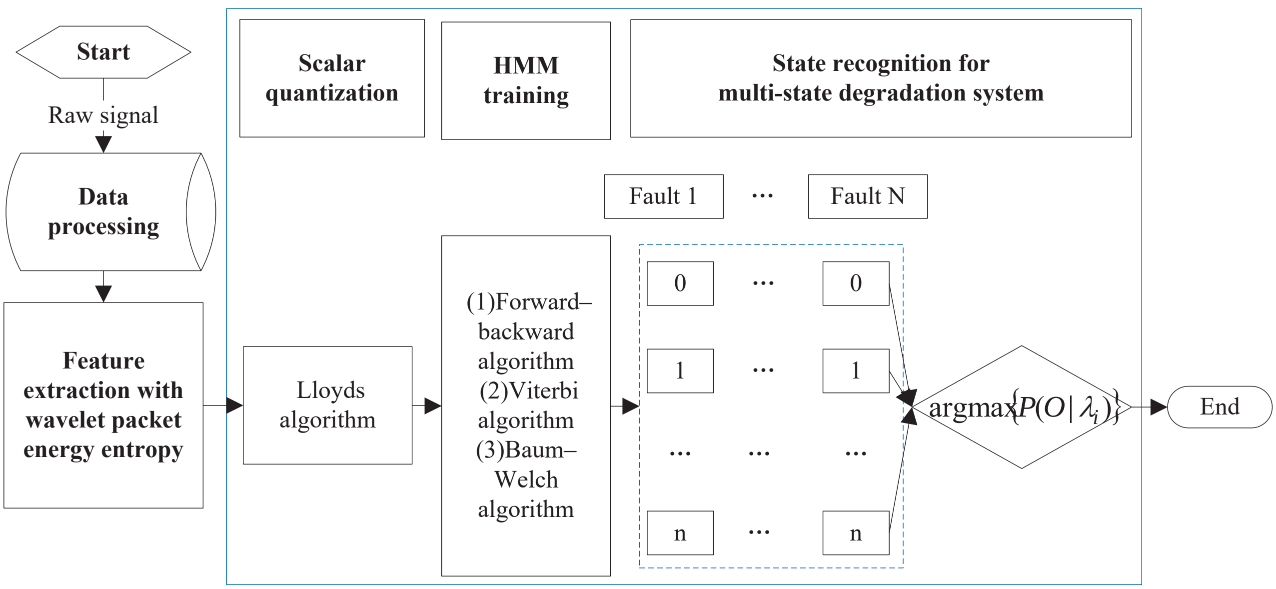


Figure 3. Dynamic diagnosis procedure of multi-state degradation system

Due to the sub-band width and frequency resolution are determined by the decomposition level j, and choosing a appropriate j can ensure to produce a good enough result. This is because a large decomposition level j may obtain a good frequency resolution but take a long time to calculate. In contrast, a small decomposition level j causes a low frequency resolution. Generally, the maximum decomposition level satisfies $j_{\max} \le [\log_2(f_s / f_m) - 1]$, and f_m is the fault characteristic frequency [23].

By the j_{th} level of decomposition, we decomposed the raw signal $x(t)$ into 2^j sub-band, and defined n sub-band as.

$$S_{j,n}(t), n = 1, 2, \cdots, 2^j \tag{23}$$

The sub-band energy of the vibration signal is defined as.

$$E_{j,n} = \int \left| S_{j,n}(t) \right|^2 dt = \sum_{k=1}^{N} \left| x_{n,k} \right|^2, \left(k = 1, 2, \cdots, N \right) \tag{24}$$

where, $x_{n,k}$ is the amplitude of $S_{j,n}(t)$, N denotes the sample number.

Each sub-band energy ratio is as follow.

$$e_{j,n} = E_{j,n} / \sum E_{j,n} (n = 1, 2, \cdots, 2^j) \text{ and } \sum_{n=1}^{n} e_{j,n} = 1 \quad (25)$$

Information entropy is the uncertainty measurement of random events, and we defined the information entropy of n sub-band as

$$H_{j,n} = -\sum_{n=1}^{n} e_{j,n} \log_2 (e_{j,n}) \quad (26)$$

After normalization processing, we obtain the normalized energy entropy of each sub-band.

$$h_n = H_{j,n} / \sum_{n=1}^{n} H_{j,n} \quad (27)$$

The feature vector for wavelet packet energy entropy is constructed as.

$$F = [h_1^T, h_2^T, \cdots, h_n^T] \quad (28)$$

(3) Scalar quantization

For a discrete HMM, the observation must be a limited discrete value, however, the feature vector extracted from vibration signal is a continuous value. Fortunately, scalar quantization is addressed the above problem, and Lloyds algorithm is an available scalar quantization method. By Lloyds algorithm, the signal is divided into $N-1$ equal portions with an increasing order, and the index value of each region is defined as.

$$index(x) = \begin{cases} 1 & x \le partition(1) \\ i & partition(i) < x \le partition(i+1) \\ N & partition(N-1) < x \end{cases} \quad (29)$$

Distortion is defined as the average of the square of the differences between the raw signal and the quantization signal.

$$distortion = \frac{1}{N} \sum_{i=1}^{M} (sig(i) - quan(i))^2 \quad (30)$$

(4) State recognition of multi-state degradation system

There are N fault modes in the multi-state degradation system and n states in each fault mode. Current state evaluation and multi-state recognition have been implemented by the proposed dynamic diagnosis approach of multi-state degradation system. Here, multi-state recognition includes fault mode classification and degradation state recognition, and the state transition matrix is defined as a equal probability.

IV. EXPERIMENTAL TESTING

The proposed dynamic diagnosis method of multi-state degradation system was verified by a bearing experimental data. As shown in Figure 4, the experimental system includes some SKF bearings, a 2*HP* driving motor (1*HP*=746*W*), a torque transducer, a dynamometer, and some control electronics. Various fault modes and fault severity of bearings were simulated in this experiment. The testing bearings with slight (fault diameters of 0.365*mm*) and serious fault (fault diameters of 0.710*mm*) were processed by an electro-discharge machining. The vertical bearing vibration signals were collected by accelerometers installed on the drive end of the motor housing with four rotation speeds (1797*rpm*, 1772*rpm*, 1750*rpm* and 1730*rpm*). As shown in table 1, 10 data samples were collected for each state under each rotation speed. Here, sample frequency, sample time and sample length are 12*kHz*, 0.1*s*, and 1200, respectively. Label F0, F1, F2 stand for normal working state, rolling and inner ring fault, respectively. Label 0, 1, 2 denote the normal, slight and serious fault state, respectively. Hence, we considered the bearing fault as a multi-state degradation system.

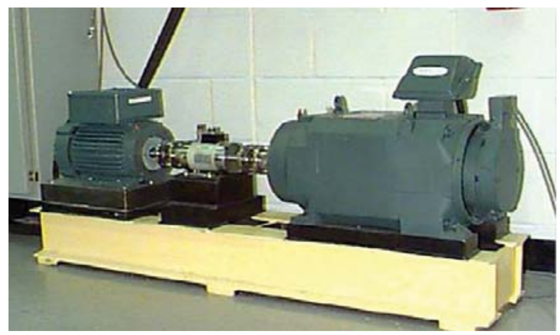


Figure 4. Bearing vibration experiment apparatus

TABLE I. THE 200 DATA SAMPLES UNDER DIFFERENT CONDITIONS

Fault	*State*	*Conditions*	*No.*
F0 No.1-40	0 No.1-40	(0,1797rpm)	1-10
		(1,1772rpm)	11-20
		(2,1750rpm)	21-30
		(3,1730rpm)	31-40
F1 No.41-120	1 No.41-80	(0,1797rpm)	41-50
		(1,1772rpm)	51-60
		(2,1750rpm)	61-70
		(3,1730rpm)	71-80
	2 No.81-120	(0,1797rpm)	81-90
		(1,1772rpm)	91-100
		(2,1750rpm)	101-110
		(3,1730rpm)	111-120
F2 No.121-200	1 No.121-160	(0,1797rpm)	121-130
		(1,1772rpm)	131-140
		(2,1750rpm)	141-150
		(3,1730rpm)	151-160
	2 No.161-200	(0,1797rpm)	161-170
		(1,1772rpm)	171-180
		(2,1750rpm)	181-190
		(3,1730rpm)	191-200

V. RESULTS AND DISCUSSIONS

A. *Description of Data Samples*

The uncertainty and the complexity of the signals can be effectively characterized by information entropy. Wavelet packet decomposition is used for signal processing, because of its high time-frequency resolution. Thus, the wavelet energy entropy includes implicit abundant fault information. Obviously, the amplitude increases with the fault severity increasing. Here, the defect frequencies of the rolling element and inner ring are 135.90~141.16Hz and 156.13~162.02Hz, respectively. Therefore, this paper selects 6 level of wavelet packet decomposition by 'daubechies5', and the width of each sub-band is 187.5 Hz. Two hundred data samples are obtained by feature extraction of wavelet packet energy entropy. Table 2 shows the training and testing data samples for current state evaluation, fault modes classification and degradation state recognition. Dataset 1 consists of the data samples of normal working state (Label F0), rolling element fault (Label F1) and inner ring fault (Label F2). Dataset 2 consists of the data samples of normal state (Label 0), slight rolling element fault (Label 1), and serious rolling element fault (Label 2).

TABLE II. DESCRIPTION OF THE TRAINING/TESTING DATA SAMPLES

Data	**Fault**	**Training/testing**	**Size**
Dataset 1	F0/F0	20/20	40
	F1/F1	20/20	40
	F2/F2	20/20	40
Data	**State**	**Training/testing**	**Size**
Dataset 2 (F1)	0/0	20/20	40
	1/1	20/20	40
	2/2	20/20	40

B. *Current State Evaluation*

When fault information is unknown, current state evaluation is validated by dataset 2. Twenty data samples of normal state (Label 0) are randomly selected for training by HMM. Sixty data samples of each state are testing, and the results are shown in Figure 5. The first 20 testing samples are in state 0, the second 20 testing samples are in state 1, the remaining 20 testing samples are in state 2. The log likelihood ratio is larger, the system is healthier. The system performance appears degradation trend, indicating that the system is deteriorated from normal to final fault. Therefore, current state evaluation has been implemented by setting the fault thresholds which are determined through continuous monitoring for each state.

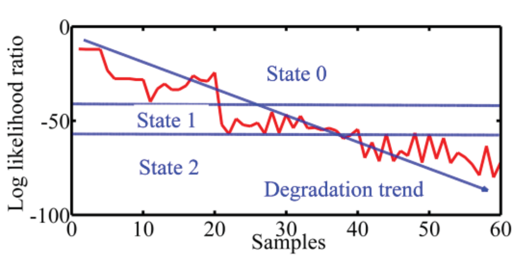


Figure 5. The results of current state evaluation

C. *Multi-State Recognition*

When fault information is determined, multi-state recognition includes fault modes classification and degradation state recognition. Fault modes classification is validated by Dataset 1, and degradation state recognition is validated by Dataset 2.

For fault modes classification, the log likelihood ratios of rolling element and inner ring fault are calculated by inputting observation sequences into the trained model, the results are shown in Figure 6. All twenty testing samples of normal state (Label F0) are classified correctly. No.20 testing sample of rolling element fault (Label F1) is misclassification. No.1, No.2, No.3, No.4 and No.5 testing samples of inner ring fault (Label F2) are misclassification. The average accuracy is more than 0.90.

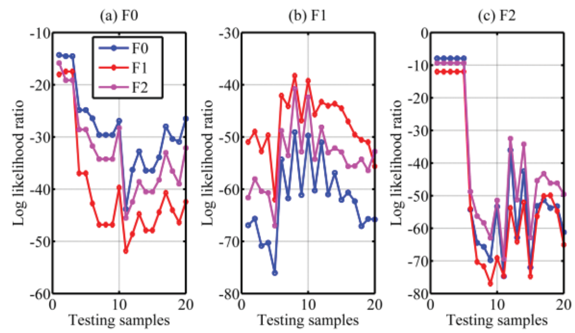


Figure 6. The results of fault modes classification

For degradation state recognition, the log likelihood ratios of rolling element slight (Label state 1) and serious (Label state 2) fault are calculated by inputting observation sequences into the trained model, the results are shown in Figure 7. All sixty testing samples are identified correctly. The average accuracy is 1.00.

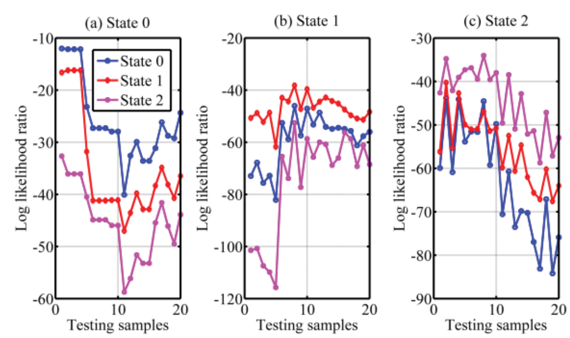
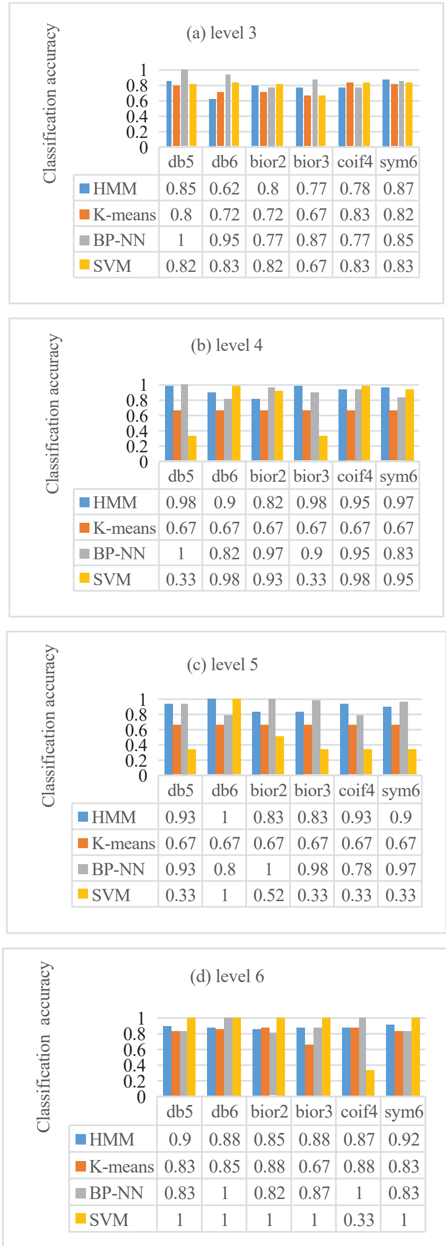


Figure 7. The results of degradation state recognition

The performance of fault modes classification and the performance of degradation state recognition are compared with *k*-means, BP neural networks (BP-NN), and Support Vector Machine (SVM). These classifiers have been proven to achieve good classification performance in fault diagnosis, pattern recognition, and so on. The effects of decomposition levels and wavelet mother functions on the classification performance are taken into account in this paper. To achieve better accuracy, the suitable values of the regularization parameters are chosen optimally for SVM classifier by grid search technique. Wavelet mother functions of 'db5' , 'db6', 'bior2.6', ' bior3.9', 'coif4', 'sym6' are selected to validate the performance of fault mode classification and the performance of degradation state recognition.



	db5	db6	bior2	bior3	coif4	sym6
HMM	0.85	0.62	0.8	0.77	0.78	0.87
K-means	0.8	0.72	0.72	0.67	0.83	0.82
BP-NN	1	0.95	0.77	0.87	0.77	0.85
SVM	0.82	0.83	0.82	0.67	0.83	0.83

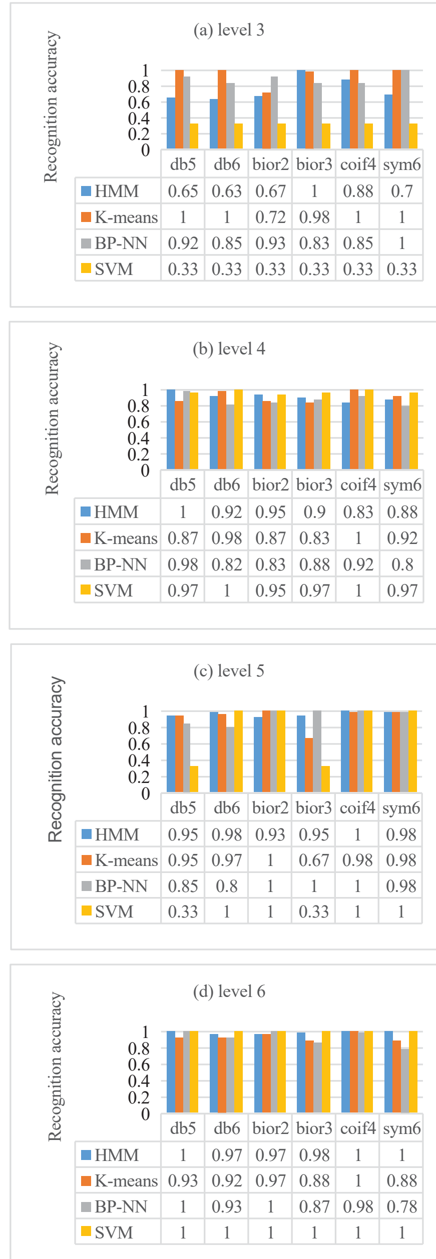
	db5	db6	bior2	bior3	coif4	sym6
HMM	0.98	0.9	0.82	0.98	0.95	0.97
K-means	0.67	0.67	0.67	0.67	0.67	0.67
BP-NN	1	0.82	0.97	0.9	0.95	0.83
SVM	0.33	0.98	0.93	0.33	0.98	0.95

	db5	db6	bior2	bior3	coif4	sym6
HMM	0.93	1	0.83	0.83	0.93	0.9
K-means	0.67	0.67	0.67	0.67	0.67	0.67
BP-NN	0.93	0.8	1	0.98	0.78	0.97
SVM	0.33	1	0.52	0.33	0.33	0.33

	db5	db6	bior2	bior3	coif4	sym6
HMM	0.9	0.88	0.85	0.88	0.87	0.92
K-means	0.83	0.85	0.88	0.67	0.88	0.83
BP-NN	0.83	1	0.82	0.87	1	0.83
SVM	1	1	1	1	0.33	1

Figure 8. Classification results of HMM, K-means, BP-NN and SVM under different wavelet families

Figure 8 and Figure 9 illustrate the summary of the classification and the recognition results of HMM, K-means, BP-NN and SVM under different types of wavelet mother functions respectively. It is found that the performance of fault modes classification and the performance of degradation state recognition are stable by HMM. The average accuracy of fault modes classification is more than 0.90, and the average accuracy of degradation state recognition is more than 0.95. However, other classifiers are severely affected by the number of decomposition level, especially SVM classifier. It is deduced that the 6 level of decomposition appears the best classification and recognition performance.



	db5	db6	bior2	bior3	coif4	sym6
HMM	0.65	0.63	0.67	1	0.88	0.7
K-means	1	1	0.72	0.98	1	1
BP-NN	0.92	0.85	0.93	0.83	0.85	1
SVM	0.33	0.33	0.33	0.33	0.33	0.33

	db5	db6	bior2	bior3	coif4	sym6
HMM	1	0.92	0.95	0.9	0.83	0.88
K-means	0.87	0.98	0.87	0.83	1	0.92
BP-NN	0.98	0.82	0.83	0.88	0.92	0.8
SVM	0.97	1	0.95	0.97	1	0.97

	db5	db6	bior2	bior3	coif4	sym6
HMM	0.95	0.98	0.93	0.95	1	0.98
K-means	0.95	0.97	1	0.67	0.98	0.98
BP-NN	0.85	0.8	1	1	1	0.98
SVM	0.33	1	1	0.33	1	1

	db5	db6	bior2	bior3	coif4	sym6
HMM	1	0.97	0.97	0.98	1	1
K-means	0.93	0.92	0.97	0.88	1	0.88
BP-NN	1	0.93	1	0.87	0.98	0.78
SVM	1	1	1	1	1	1

Figure 9. Recognition results of HMM, K-means, BP-NN and SVM under different wavelet families

For the 6 level of decomposition of the raw signals, it can be inferred that HMM and SVM classifiers have achieved good classification and identification performance. But the average classification accuracy of 0.33 is achieved by using SVM classifier for 'coif4'. The HMM classifier will obtain a better accuracy by using different wavelet mother functions, which is more stable than SVM classifier. It is deduced that 'db5' appears the best performance.

From above results, HMM classifier shows more stable and better performance in fault modes classification and degradation state recognition. Comparing with the classifiers of *K*-means, BP-NN and SVM, HMM classifier suffers the smallest effect.

VI. Conclusions

A dynamic diagnosis method of multi-state degradation system was proposed to develop current state evaluation, fault modes classification, and degradation state recognition in this paper. The accuracy and efficiency of dynamic diagnosis approach was tested by bearing experiment.

(1) HMM classifier had been successfully employed in incomplete information system and complete information system. Current health state evaluation had been implemented when fault information is unknown. While fault information is determined, fault modes classification and degradation state recognition have been carried out.

(2) Compared with three classifiers of *K*-means, BP-NN, SVM, HMM classifier suffered the smallest effects of the number of decomposition level and wavelet mother functions. HMM had achieved good classification and identification performance and provided model interpretation. Results showed that HMM was the most suitable for dynamic diagnosis in multi-state degradation system.

Acknowledgment

This research sincerely thanks bearing data center of Western Reserve University in the United States for the bearing experiment data (http://www.eees.ease.edu/laboratory/bearing). This study was supported by National Natural Science Foundation of China (Grant no. 61463021 and 61963018), Natural Science Foundation of Jiangxi Province in China (Grant no. 20181BAB202020), Science and Technology Project in Education Department of Jiangxi Province in China (Grant no. GJJ14420 and GJJ180494), and Doctoral Scientific Research Foundation of Jiangxi University of Science and Technology in China (Grant no. 3401223356).

References

[1] M. S. Crouse, R. D. Nowak and R. G. Baraniuk, “Wavelet-based statistical signal processing using hidden Markov models,” IEEE Transactions on Signal Processing, vol. 46, pp. 886-902, April 1998.

[2] A. Cichocki, S. C. Douglas and S. Amari, “Robust techniques for independent component analysis (ICA) with noisy data,” Neurocomputing, vol. 22, pp. 113-129, November 1998.

[3] S. Huang, D.Yang, Y. Ge and X. Zhang, “Combined supervised information with PCA via discriminative component selection,” Information Processing Letters, vol. 115, pp. 812-816, June 2015.

[4] A. Folch-Fortuny, F. Arteaga and A. Ferrer, “PCA model building with missing data: New proposals and a comparative study,” Chemometrics & Intelligent Laboratory Systems, vol. 146, pp. 77-88, August 2015.

[5] R. W. Swiniarski and A. Skowron, “Rough set methods in feature selection and recognition,” Pattern Recognition Letters, vol. 24, pp. 833-849, March 2003.

[6] R. A. Rios and R. F. D. Mello, “Applying Empirical Mode Decomposition and mutual information to separate stochastic and deterministic influences embedded in signals,” Signal Processing, vol. 118, pp. 159-176, July 2016.

[7] P. Nguyen, M. Kang, J. M. Kim, B. H. Ahn, J. M. Ha and B. K. Choi, “Robust condition monitoring of rolling element bearings using de-noising and envelope analysis with signal decomposition techniques,” Expert Systems with Applications, vol. 42, pp. 9024-9032, December 2015.

[8] A. Rai, and S. H. Upadhyay, “A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings,” Tribology International, vol. 96, pp. 289-306, April 2016.

[9] I. El-Thalji and E. Jantunen, “A summary of fault modelling and predictive health monitoring of rolling element bearings,” Mechanical Systems & Signal Processing, vol. 60, pp. 252-272, August 2015.

[10] B . Zhang, S. Zhang, and W. Li, “Bearing performance degradation assessment using long short-term memory recurrent network,” Computers in Industry, vol. 106, pp. 14-29, April 2019.

[11] G. Ni, J. Chen, and H. Wang, “Degradation assessment of rolling bearing towards safety based on random matrix single ring machine learning,” Safety Science, vol. 118, pp. 403-408, October 2019.

[12] Z. Li, Y. Wang, and K. Wang. “A deep learning driven method for fault classification and degradation assessment in mechanical equipment,” Computers in Industry, vol. 104, pp. 1-10, January 2019.

[13] W. Caesarendra, A. Widodo and B. S. Yang, “Application of relevance vector machine and logistic regression for machine degradation assessment,” Mechanical Systems & Signal Processing, vol. 24, pp. 1161-1171, May 2010.

[14] X. Zhu, Y. Zhang and Y. Zhu, “Bearing performance degradation assessment based on the rough support vector data description,” Mechanical Systems & Signal Processing, vol. 34, pp. 203-217, January 2013.

[15] J. Yu, “Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models,” Mechanical Systems and Signal Processing, vol. 25, pp. 2573-2588, October 2011.

[16] J. Li, X. Zhang, X. Zhou and L. Lu, “Reliability assessment of wind turbine bearing based on the degradation-Hidden-Markov model,” Renewable energy, vol. 132, pp. 1076-1087, March 2019.

[17] H. T. Pham, B. S. Yang and T. T. Nguyen, “Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine,” Mechanical Systems and Signal Processing, vol. 32, pp. 320-330, October 2012.

[18] M. Dong, D. He, P. Banerjee and J. Keller, “Equipment health diagnosis and prognosis using hidden semi-Markov models,” International Journal of Advanced Manufacturing Technology, vol. 30, pp. 738-749, October 2006.

[19] G. Qiu, S. Huang, L. Zhu, X. Su and Y. Chen, “Risk assessment of multi-state bayesian network in an oil gathering and transferring system,” Procedia Engineering, vol. 130, pp. 1514-1523, September 2015.

[20] M. Mohamed and P. Gader, “Generalized hidden Markov models I. Theoretical frameworks,” IEEE Transactions on Fuzzy Systems, vol. 8, pp. 67-81, February 2000.

[21] L. R. Rabiner and B. H. Juang, “An introduction to hidden Markov models,” IEEE ASSP Magazine, vol. 3, pp. 4-16, January 1986.

[22] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proceedings of the IEEE, vol. 77, pp. 257-286, February 1989.

[23] X. Fan, M. Zuo, “Gearbox fault detection using Hilbert and wavelet packet transform,” Mechanical Systems and Signal Processing, vol. 20, pp. 966-982, May 2006.