Front. Mech. Eng. 2018, 13(2): 301–310
https://doi.org/10.1007/s11465-017-0449-7

RESEARCH ARTICLE

Yanfeng PENG, Junsheng CHENG, Yanfei LIU, Xuejun LI, Zhihua PENG

An adaptive data-driven method for accurate prediction of remaining useful life of rolling bearings

Abstract A novel data-driven method based on Gaussian mixture model (GMM) and distance evaluation technique (DET) is proposed to predict the remaining useful life (RUL) of rolling bearings. The data sets are clustered by GMM to divide all data sets into several health states adaptively and reasonably. The number of clusters is determined by the minimum description length principle. Thus, either the health state of the data sets or the number of the states is obtained automatically. Meanwhile, the abnormal data sets can be recognized during the clustering process and removed from the training data sets. After obtaining the health states, appropriate features are selected by DET for increasing the classification and prediction accuracy. In the prediction process, each vibration signal is decomposed into several components by empirical mode decomposition. Some common statistical parameters of the components are calculated first and then the features are clustered using GMM to divide the data sets into several health states and remove the abnormal data sets. Thereafter, appropriate statistical parameters of the generated components are selected using DET. Finally, least squares support vector machine is utilized to predict the RUL of rolling bearings. Experimental results indicate that the proposed method reliably predicts the RUL of rolling bearings.

Keywords Gaussian mixture model, distance evaluation technique, health state, remaining useful life, rolling bearing

Received November 16, 2016; accepted March 8, 2017

Yanfeng PENG, Junsheng CHENG (✉)
State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, China
E-mail: chengjunsheng1@163.com, 515667195@qq.com

Yanfeng PENG, Junsheng CHENG
College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China

Yanfeng PENG, Yanfei LIU, Xuejun LI
Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Xiangtan 411201, China

Zhihua PENG
School of Mathematics and Physics, University of South China, Hengyang 421001, China

1 Introduction

Bearing fault is a major source of failures in mechanical drive systems; thus, predicting the remaining useful life (RUL) of rolling bearings has been a key research topic in condition-based maintenance [1–12]. The methods for the prognosis of rolling bearing RUL can be classified into two categories: Model-based methods [1–4] and data-driven methods [5–12]. In model-based methods, the RUL is estimated using physical laws or by solving several deterministic equations obtained from empirical data. An approach based on a finite element model and a bearing spall propagation model was proposed by Marble and Morton [1] for health state prediction of propulsion system bearings. Liao et al. [2] developed an RUL prediction method based on logistic and hazard regression models. Tian and Liao [3] proposed a proportional hazard model to predict the RUL of rolling bearings. Model-based methods can significantly improve the accuracy of RUL prediction if properly used. However, damage propagation processes and equipment dynamic response are complex, thereby bringing difficulty in building authentic models.

Data-driven methods are based on the data derived from various conditions (e.g., vibration, temperature, and pressure) and signal processing techniques. Compared with model-based methods, data-driven methods can be applied to more complex situations because of ignoring the physical laws and equations of rolling bearings. Gebraeel et al. [5] proposed an approach based on artificial neural network to predict the RUL of bearings. Di Maio et al. [6] developed a method based on relevance vector machine to estimate the RUL of bearings. Ben Ali et al. [7] proposed an approach based on Weibull distribution and artificial neural network. However, existing data-driven methods for RUL prediction have no unified standards for estimating the health state of rolling bearings. The health states are determined empirically by qualified experts in experiment [7,11].

Gaussian mixture model (GMM) has been widely applied in many areas, such as clinical decision [13], computer vision, [14], and mechanical fault diagnosis [8]. Considerable research has shown that GMM is suitable for diagnosing bearing vibration signals [15,16], but most previous works only focused on quantifying and modeling of bearing degradation performance. In the current study, GMM is used to cluster the extracted features of vibration signals. The maximum likelihood estimation of GMM is performed by expectation maximization (EM) [15,16], and the number of clusters is determined by the minimum description length (MDL) principle [13]. Thus, either the health state of the data sets or the number of the states is obtained automatically. Meanwhile, the abnormal data sets can be recognized during the clustering process and can be removed from the training data sets.

Not all features can positively contribute to the prediction of rolling bearing RUL because some of them may lower the identification accuracy and the computational efficiency. Thus, appropriate features (i.e., salient features) should be selected before they are inputted into a classifier. Distance evaluation technique (DET) [17–19] is thus applied to obtain salient features in this study. The main idea of DET is to select salient features that show low intra-class variations and high intra-class variations by use of effectiveness factor [18]. Least squares support vector machine (LS-SVM) is widely used in the fault diagnosis of rolling bearings and shows a satisfactory performance [11]. Thus, this method is used to estimate the rolling bearing RUL in this study.

A novel adaptive method using GMM [14,20] and DET [17–19] is proposed in this study to improve the adaptability and validity of the current approaches. The rest of the paper is organized as follows. The process of the proposed method is presented in Section 2. The basic theory of the relevant methods is also introduced in this section. The experimental effectiveness of the method is analyzed in Section 3. The conclusions and discussion are provided in Section 4.

2 Methodology

2.1 Feature extraction

The features of the vibration signals should be extracted prior to the clustering process. Empirical mode decomposition (EMD) is a commonly used time-frequency analysis method and is widely applied as a processing approach for the feature extraction of mechanical vibration signal diagnosis [21,22]; thus, the method is used to decompose the data sets into a number of IMFs (intrinsic mode functions) in the proposed method. Additional details of EMD can be found in Refs. [21–24]. Given that each IMF represents the natural oscillatory mode of the original signal, features extracted from the components are more effective than the features extracted from the original signal for bearing vibration signal diagnosis in some cases. The statistical parameters from the original vibration signals and the IMFs generated by EMD are extracted to obtain accurate physical features. For calculation convenience, all the statistical parameters are normalized in the proposed method. The common frequency domain and time domain statistical parameters applied in this study are shown as below:

Frequency domain statistical parameters:

1) Mean value:

$$F_1=\frac{1}{N}\sum_{n=1}^{N}x(n).$$

2) Root mean square:

$$F_2=\sqrt{\frac{1}{N}\sum_{n=1}^{N}x^2(n)}.$$

3) Square root amplitude:

$$F_3=\left(\frac{1}{N}\sum_{n=1}^{N}\sqrt{|x(n)|}\right)^2.$$

4) Mean amplitude:

$$F_4=\frac{1}{N}\sum_{n=1}^{N}|x(n)|.$$

5) Maximum peak:

$$F_5=\frac{1}{2}(\max(x(n))-\min(x(n))).$$

6) Standard deviation:

$$F_6=\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left(x(n)-F_1\right)^2}.$$

7) Skewness:

$$F_7=\frac{1}{N-1}\sum_{n-1}^{N}\left(\frac{x(n)-F_1}{F_6}\right)^3.$$

8) Kurtosis:

$$F_8=\frac{1}{N-1}\sum_{n-1}^{N}\left(\frac{x(n)-F_1}{F_6}\right)^4.$$

9) Crest factor:

$$F_9 = \frac{F_5}{F_2}.$$

10) Clearance factor:

$$F_{10} = \frac{F_5}{F_3}.$$

11) Shape factor:

$$F_{11} = \frac{F_2}{F_4}.$$

12) Crest factor:

$$F_{12} = \frac{F_5}{F_4}.$$

Time domain statistical parameters:
13) Spectral amplitude mean value:

$$F_{13} = \frac{1}{M}\sum_{k=1}^{M} s(k).$$

14) Spectral amplitude standard deviation:

$$F_{14} = \sqrt{\frac{1}{M-1}\sum_{k=1}^{M}(s(k)-F_{13})^2}.$$

15) Spectral amplitude skewness:

$$F_{15} = \frac{1}{M-1}\sum_{k=1}^{M}\left(\frac{s(k)-F_{13}}{F_{14}}\right)^3.$$

16) Spectral amplitude kurtosis:

$$F_{16} = \frac{1}{M-1}\sum_{k=1}^{M}\left(\frac{s(k)-F_{13}}{F_{14}}\right)^4.$$

17) Spectral gravity frequency:

$$F_{17} = \frac{\sum_{k=1}^{M} f(k)s(k)}{\sum_{k=1}^{M} s(k)}.$$

18) Spectral root mean square frequency:

$$F_{18} = \sqrt{\frac{\sum_{k=1}^{M} f^2(k)s(k)}{\sum_{k=1}^{M} s(k)}}.$$

19) Spectral root 4/2-moment ratio:

$$F_{19} = \sqrt{\frac{\sum_{k=1}^{M} f^4(k)s(k)}{\sum_{k=1}^{M} f^2(k)s(k)}}.$$

20) Spectral standard deviation frequency:

$$F_{20} = \sqrt{\frac{\sum_{k=1}^{M} (f(k)-F_{17})^2 s(k)}{\sum_{k=1}^{M} s(k)}}.$$

21) Spectral frequency skewness:

$$F_{21} = \sum_{k=1}^{M}\frac{\left(\frac{f(k)-F_{17}}{F_{20}}\right)^3 s(k)}{\sum_{j=1}^{M} s(j)}.$$

22) Spectral frequency kurtosis:

$$F_{21} = \sum_{k=1}^{M}\frac{\left(\frac{f(k)-F_{17}}{F_{20}}\right)^4 s(k)}{\sum_{j=1}^{M} s(j)}.$$

The IMFs covering high-frequency bands are usually generated first in the procedure of EMD. Furthermore, bearing fault often induces high-frequency bands [18]. Therefore, the first several IMFs generated by EMD are selected in practical applications. Considering the above-mentioned fact, only the first three IMFs are selected for extracting statistical features.

2.2 Health state clustering based on GMM

GMM is the sum of the weight of several Gaussian components. The density of a specific random variable is represented by the Gaussian components [14,20]. In GMM, a group of Gaussian functions allow an improved modeling capability while each Gaussian function has its co-variance and mean matrix. GMM composed of three Gaussian components is shown in Fig. 1. The mathematical formula of GMM is as follows [14,20]:

$$P(\boldsymbol{x}|\boldsymbol{\Theta}) = \sum_{i=1}^{M} w_i P\left(\boldsymbol{x}|\mu_i, \Sigma_i\right), \tag{1}$$

where M is the largest number of the Gaussian components, and $\boldsymbol{x}$ is the random variable, $\boldsymbol{\Theta}$ are the GMM parameters including the covariance matrix $\sum$, the mean vector $\boldsymbol{\mu}$, and the weight w ($w\geqslant 0$, $\sum_{i=1}^{M} w_i = 1$), $P(\boldsymbol{x}|\mu_i, \sum_i)$ is the density function of the ith component, which can be described as follows:

$$P\left(\boldsymbol{x}|\mu_i, \sum\nolimits_i\right) = \frac{1}{(2\pi)^{\frac{D}{2}}|\sum_i|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(x-\mu_i)^{\mathrm{T}}\sum_i^{-1}(x-\mu_i)}, \tag{2}$$

where D represents the dimension of the vector $\boldsymbol{x}$.

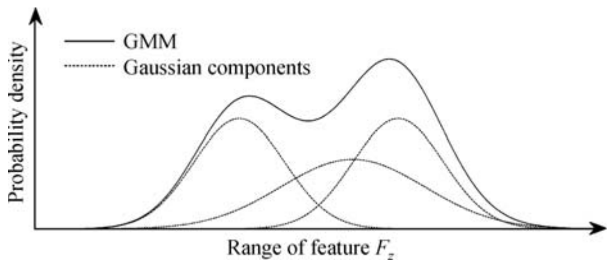

Fig. 1 Illustration of GMM

EM [15,16] is the most common method for the estimation of the parameters of GMM. Supposing the number of data sets is N, the maximum likelihood of the parameter $\boldsymbol{\Theta}$ is obtained during the iterative process of EM by optimizing the following likelihood function:

$$l(\boldsymbol{\Theta}) = \log\prod\nolimits_{t=1}^{N} p\left(\boldsymbol{x}^{(t)}|\boldsymbol{\Theta}\right) = \sum_{t=1}^{N} \log p\left(\boldsymbol{x}^{(t)}|\boldsymbol{\Theta}\right). \tag{3}$$

The optimization process involves two steps. The parameters of $\boldsymbol{\Theta}$ are initialized first and then they are updated iteratively until convergence. The parameters are updated as follows in each iteration j:

$$w_i^{(j+1)} = \frac{\sum_{t=1}^{N} h_i^{(j)}(t)}{N}, \tag{4}$$

$$\mu_i^{j+1} = \frac{\sum_{t=1}^{N} h_i^{(j)}(t)x^{(t)}}{\sum_{t=1}^{N} h_i^{(j)}(t)}, \tag{5}$$

$$\sum\nolimits_i^{j+1} = \frac{\sum_{t=1}^{N} h_i^{(j)}(t)\left[x^{(t)} - \mu_i^{(j+1)}\right]\left[x^{(t)} - \mu_i^{j+1}\right]^{\mathrm{T}}}{\sum_{t=1}^{N} h_i^{(j)}(t)x^{(t)}}, \tag{6}$$

where $h_i^{(j)}$ is the posterior probability of the ith component at the jth iteration, and its equation is as follows:

$$h_i^{(j)}(t) = \frac{w_i^{(j)}P\left(x^{(t)}|\mu_i^{(j)}, \sum_i^{(j)}\right)}{\sum_{i=1}^{M} w_i^{(j)}P\left(x^{(t)}|\mu_i^{(j)}, \sum_i^{(j)}\right)}. \tag{7}$$

The number of the Gaussian components should be identified first prior to adaptively clustering the health states. In recent years, many approaches, such as MDL principle, Akaike's information criterion, and Laplace empirical criterion [14], have been proposed to attain the said goal. MDL is a classic algorithm and is used to select the number of clusters, and the experimental results show a satisfactory performance. MDL is described as follows:

$$MDL(K, \Theta) = -\sum_{t=1}^{N} \log\left(\sum_{j=1}^{K} p(x^{(t,j)}|\Theta)\right) + \frac{1}{2}L\log(NR), \tag{8}$$

where K is the clustering number, R is the number of the selected features, and L is defined as follows:

$$L = K\left(1 + R + \frac{(R+1)R}{2}\right) - 1. \tag{9}$$

When the MDL principle is applied to obtain the number of clusters, a maximum number of cluster $K_{\max}$ should be set first. Then, the data sets are clustered by GMM when $K = 1, 2, ..., K_{\max}$. The MDL of each clustering result should be calculated, and the result corresponding to the minimum MDL value is supposed to be the best.

2.3 Feature selection based on DET

The selection of DET [17–19] to obtain salient features in this study has been explained in the introduction section. As mentioned earlier, the main idea of DET is to select salient features that show low intra-class variations and high intra-class variations using effectiveness factor.

Assuming $F_{i,j,k}$ ($i = 1, 2, ..., C; j = 1, 2, ..., J; k = 1, 2, ..., N_i$) is the jth statistical parameter of the kth sample with ith health state. C, J, and N_i are the number of conditions, statistical parameters and samples, respectively. The process of DET is as follows:

1) The average distance of the samples with the same condition is calculated as

$$d_{i,j} = \frac{1}{N_i(N_i - 1)}\sum_{k,l=1}^{N_i} |F_{i,j,k} - F_{i,j,l}|$$
$$(k\neq l;\ i = 1, 2, ..., C;\ j = 1, 2, ..., J), \tag{10}$$

and then the average distance of all conditions is obtained as

$$d_j^W = \frac{1}{C}\sum_{i=1}^{C} d_{i,j}\ \ (j = 1, 2, ..., J). \tag{11}$$

2) The average value of each parameter of all samples with the same condition is calculated as

$$u_{i,j} = \frac{1}{N_i}\sum_{k=1}^{N_i} F_{i,j,k}\ \ (i = 1, 2, ..., C;\ j = 1, 2, ..., J), \tag{12}$$

and then the average distance between the average values of parameters with different conditions is obtained as

$$d_j^b = \frac{1}{C(C-1)} \sum_{i,m=1}^{C} |u_{i,j} - u_{m,j}| \ (i \neq m;\ j = 1, 2, ..., J). \tag{13}$$

3) The effectiveness factor is calculated as

$$\alpha_j = \frac{d_j^b}{d_j^W} \ (j = 1, 2, ..., J), \tag{14}$$

and then the effectiveness factor is normalized by its maximum value as

$$\alpha_j' = \frac{\alpha_j}{\max(\alpha_j)} \ (j = 1, 2, ..., J),\ \alpha_j' \in [0, 1]. \tag{15}$$

Large value of normalized effectiveness factor α_j' implies that feature p_j can be used to effectively identify different conditions. The first several features with large α_j' are chosen to be the salient features.

The salient features should be normalized as the value ranges of the salient features vary.

$$f_{i,j}' = \frac{f_{i,j}}{\max\limits_{i=1,2,...,n}(|f_{i,j}|)},\ j = 1, 2, ..., J',\ f_{i,j}' \in [-1, 1], \tag{16}$$

where i and j indicate that $f_{i,j}$ is the jth feature calculated from the ith sample, n is the number of the training samples, and J' is the number of selected features.

2.4 Proposed method

In this study, a novel method based on GMM and DET is proposed for predicting the RUL of rolling bearings. The flowchart of the proposed method is shown in Fig. 2. The main procedures of the method are as follows:

1) Decompose all the data sets into several IMFs by use of EMD in consideration of the first M components.

2) Calculate the 22 statistical parameters of the original data sets and the IMFs of the data sets generated by EMD. Given that the first M IMFs are considered, the number of the constructed feature vectors F_j is $(M+1) \times 22$. All the statistical parameters are normalized in the proposed method.

3) Obtain the health states of rolling bearings by clustering the feature vectors with GMM. The number of the health states is determined by the MDL principle. Then, remove the abnormal signals from training data sets.

4) After identifying the health states, select the salient features by DET. Calculate the normalized effectiveness factor $\alpha_j'\ (j = 1, 2, ..., (M+1) \times 22)$ of every feature vector. If the α_j' of a feature vector is larger than the selected threshold value, then the corresponding feature is chosen to be a salient feature.

5) Train and test the LS-SVM classifier with training and testing data sets, and then output the RUL results.

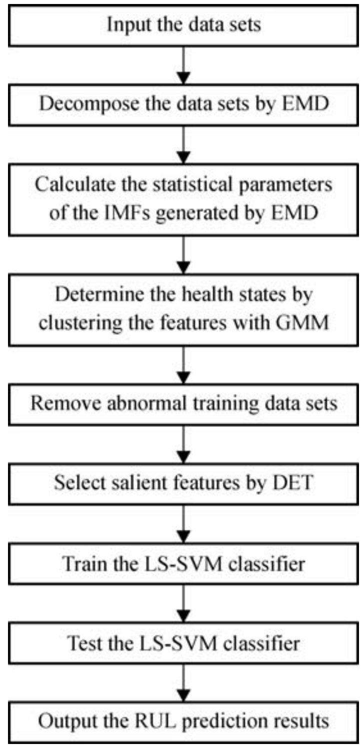

Fig. 2 Flowchart of the proposed method

3 Experimental analysis

The proposed method is evaluated by applying the data sets provided by the Center for Intelligent Maintenance Systems at the University of Cincinnati [25–27]. Rexnord ZA-2115 bearings are used on the test rig. A total of 16 rollers are contained in each row of bearings. PCB 353B33 High Sensitivity Quartz ICP accelerometers are utilized on the bearing housing. The test rig of the experiment is shown in Fig. 3. Four rolling bearings are tested in the experiment. The rotation speed is 2000 r/min, and the radial load is 6000 lbs (1 lbs = 0.4535 kg). The test lasted for 35 days until the bearings fail in obtaining run-to-failure data sets of bearings.

The vibration signals are recorded every 10 min. A total of 20480 data points are contained in each file, and the sampling rate is 20 kHz. Three tests are implemented with all the four rolling bearings in each test, thereby obtaining 12 data sets. Only three bearings fail in the experiment, and thus, their corresponding data sets are utilized. The three data sets are denoted as data sets 1, 2, and 3. The vibration signals of the three data sets are shown in Fig. 4. Additional information about the three data sets is provided in Table 1. One-third of the individual files in the data sets with the same interval are selected to be the testing data sets, and the others are considered the training data sets.

Features of the vibration signals should be extracted

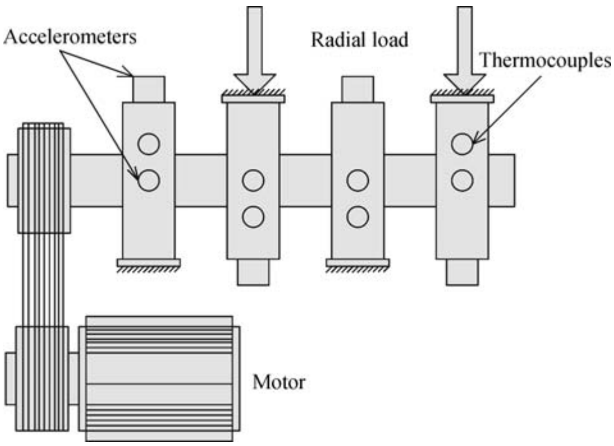

Fig. 3 Bearing test rig

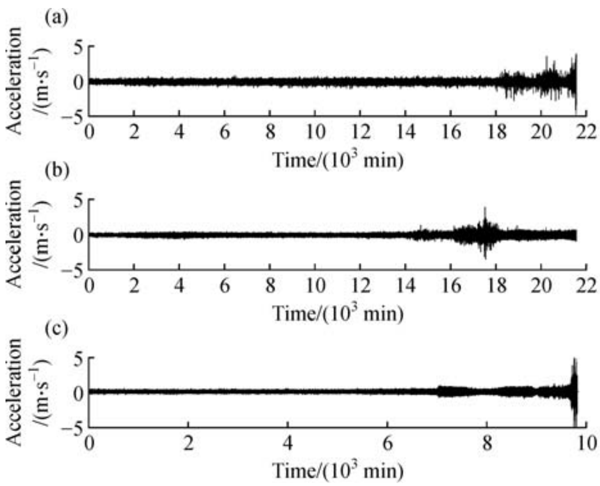

Fig. 4 Vibration signals of three data sets. (a) Data set 1; (b) data set 2; (c) data set 3

prior to the clustering process. Thus, all the individual files of the data sets are decomposed by EMD first. Then, the features are extracted by calculating the 22 statistical parameters. Considering that only the first three IMFs and the original signals are considered, 22 features of the original signals and 22×3 features of the IMFs are obtained by EMD. Thus, combined with the features of the original signals, 22×4 (i.e., 88) features of each data set are obtained for testing the effectiveness of the proposed method.

Each data set is clustered into several health states by GMM with the calculated statistical parameters. The classification of the health states generated by GMM and the skewnesses of the three data sets are presented in Figs. 5–7 for intuitively describing the clustering results. Figures 5–7 show that the health states are closely related to the fluctuation of the features. Furthermore, the numbers of the abnormal data identified by GMM in data sets 1, 2, and 3 are 76, 81, and 8, respectively. Thus, the abnormal data can be removed from the training data sets prior to the LS-SVM classification and prediction process. The

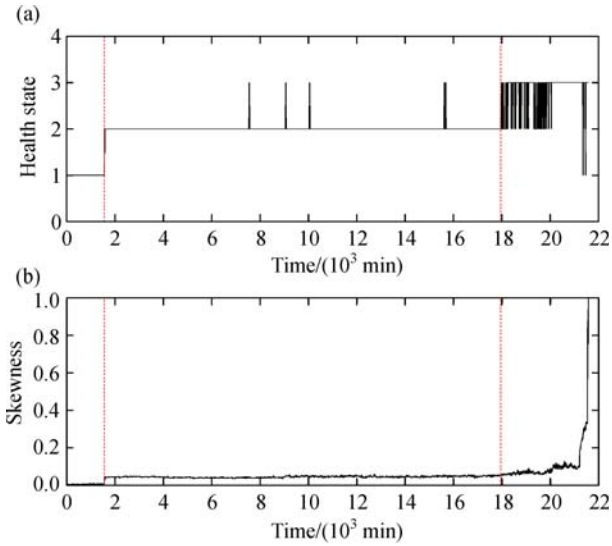

Fig. 5 Clustering results of data set 1. (a) Clustered health state; (b) skewness

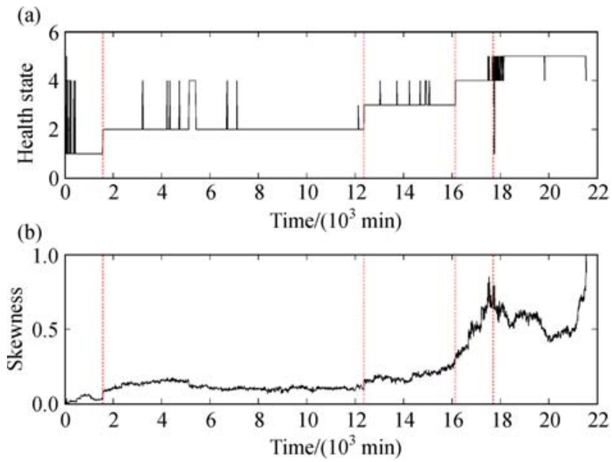

Fig. 6 Clustering results of data set 2. (a) Clustered health state; (b) skewness

Table 1 Information of the experimental data sets

Data set	Test	Bearing	Break time/min	Degradation type	Maximum magnitude/(m·s^{-2})
1	1	3	21560	Inner race	5
2	1	4	21560	Roller	4
3	2	1	9840	Outer race	5

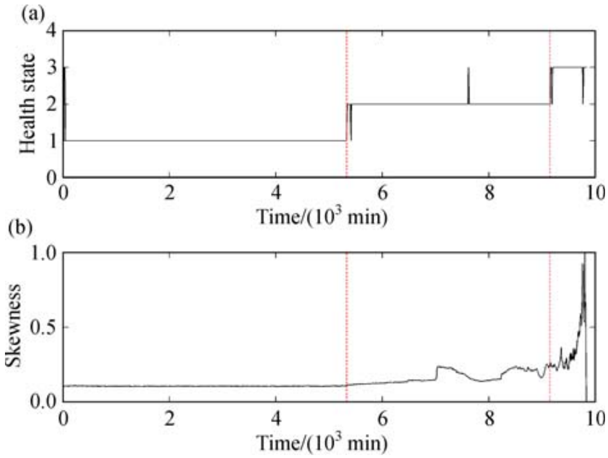

Fig. 7 Clustering results of data set 3. (a) Clustered health state; (b) skewness

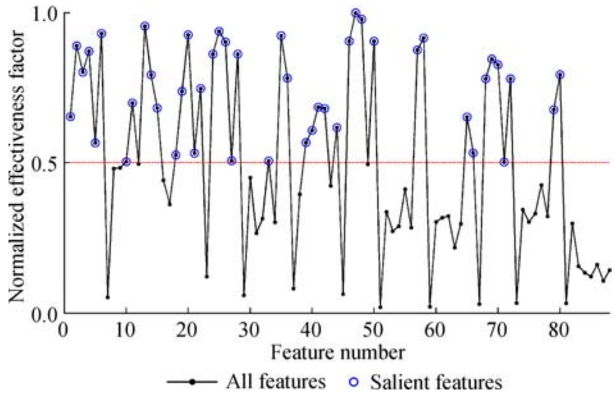

Fig. 9 Salient features of data set 2

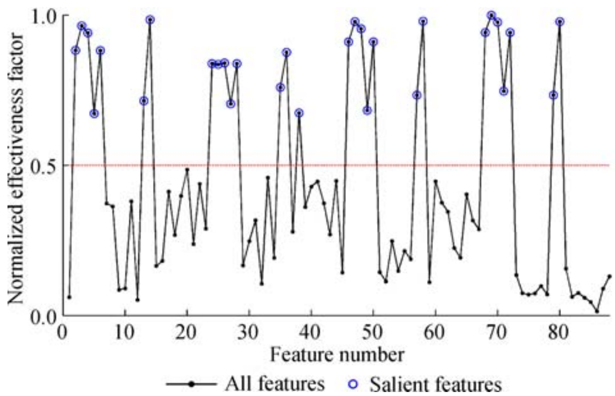

Fig. 10 Salient features of data set 3

numbers of the data removed from the training data sets of data sets 1, 2, and 3 are 50, 56, and 6, respectively. The abnormal data in the testing data sets are retained because determining whether the testing vibration signal is abnormal beforehand is difficult in practical application.

After determining the health states, DET is applied to select salient features in accordance with the health states. The threshold value is set to be 0.5 in this study. If the effectiveness factor α_j' of a feature vector is larger than 0.5, then the corresponding feature is chosen to be a salient feature. The feature selection results are shown in Figs. 8–10. A total of 14, 44, and 29 salient features are selected from the 88 features of data sets 1, 2, and 3, respectively.

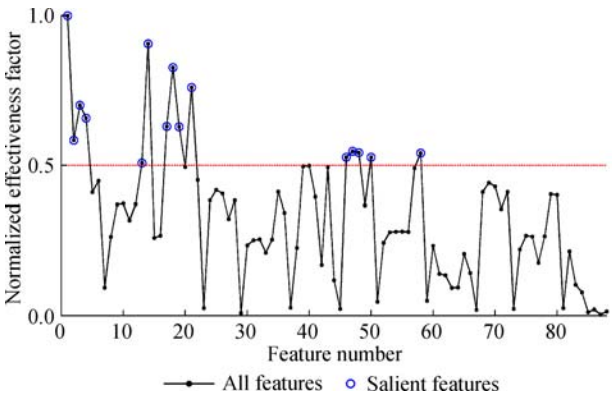

Fig. 8 Salient features of data set 1

The validity of the proposed method is verified using four methods: The proposed method (marked as Method 1), the proposed method without GMM clustering (uniformly divided into three health states and marked as Method 2), the proposed method without using DET to select feature (marked as Method 3), and the proposed method without removing the abnormal data from the training data sets (marked as Method 4). The prediction and classification results of the rolling bearing RUL are generated by LS-SVM.

Figures 11–13 show the RUL prediction results (LP: Life percentage) of the three data sets. The predicted results of Methods 1 and 2 are obviously more approximate to the real values compared with the two other methods. Meanwhile, the fluctuations of the curves generated by Methods 1 and 2 are small. Several parameters generated from the classification and RUL prediction results of the three data sets are presented in Tables 2–4 for further demonstrating the effectiveness of the proposed method. CA_i is the mean classification accuracy of the ith health state; CA is the mean classification accuracy of all the health states; EE and CC are the energy error and the correlation coefficient between the predicted RUL and the real values, respectively. EE and CC are defined in Eqs. (17) and (18).

$$EE = \frac{\sum_{n=1}^{N} (S_i(n) - I_i(n))^2}{\sum_{n=1}^{N} I_i(n)^2}, \tag{17}$$

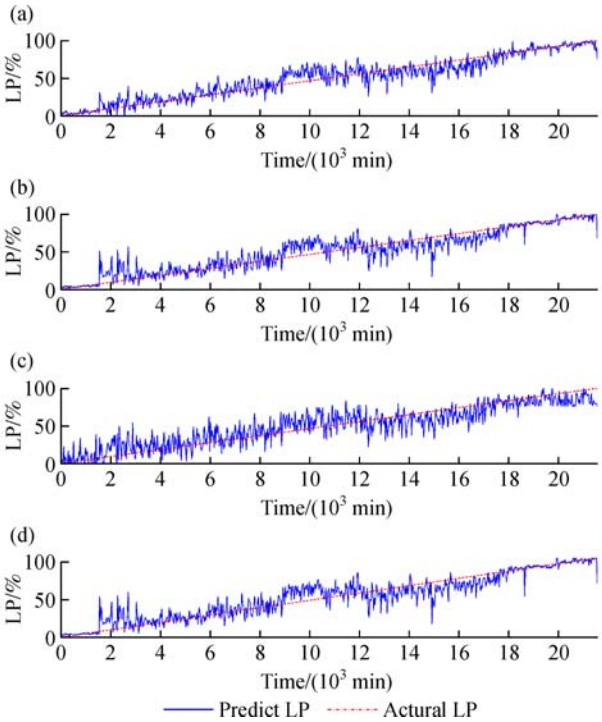

Fig. 11 RUL prediction results of data set 1. (a) Method 1; (b) Method 2; (c) Method 3; (d) Method 4

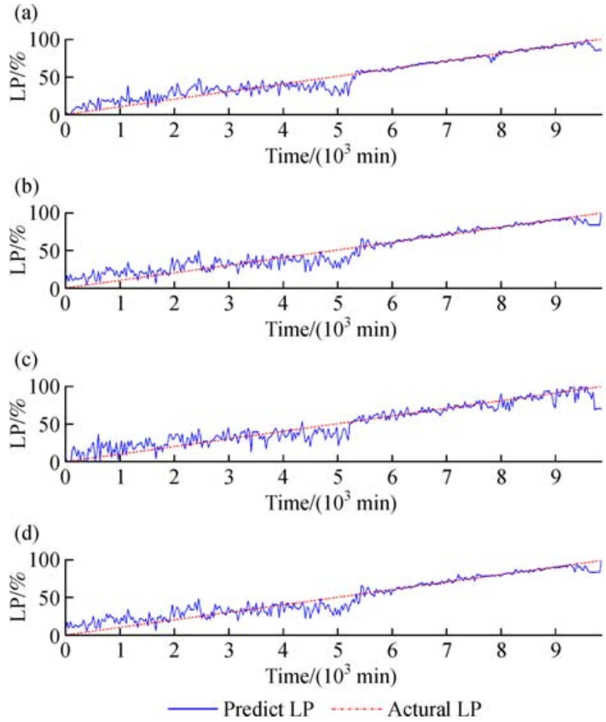

Fig. 13 RUL prediction results of data set 3. (a) Method 1; (b) Method 2; (c) Method 3; (d) Method 4

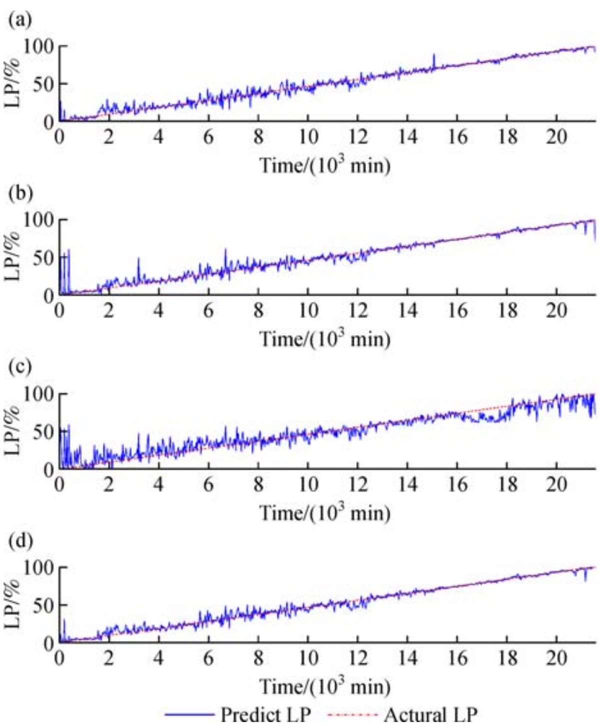

Fig. 12 RUL prediction results of data set 2. (a) Method 1; (b) Method 2; (c) Method 3; (d) Method 4

$$CC = \frac{\sum_{n=1}^{N}\left(S_i(n)-\overline{S}_i\right)\left(I_i(n)-\overline{I}_i\right)}{\sqrt{\sum_{n=1}^{N}\left(S_i(n)-\overline{S}_i\right)^2\sum_{n=1}^{N}\left(I_i(n)-\overline{I}_i\right)^2}}, \tag{18}$$

where N is the number of points in each data set, $I_i(t)$ and $S_i(t)$ are the real values and the predicted RUL, respectively, $\overline{I}_i$ and $\overline{S}_i$ are the means of $I_i(t)$ and $S_i(t)$, respectively. Tables 2–4 show that the classification accuracy and the RUL prediction accuracy generated by the proposed method are superior to those by the three other methods. Meanwhile, the proposed method shows a satisfactory performance in the experimental analysis. Therefore, the proposed method is effective in classifying health states and predicting rolling bearing RUL.

Table 2 Experimental results of data set 1

Method	CA_1	CA_2	CA_3	CA	EE	CC
1	1.0000	1.0000	0.6583	0.9429	0.0258	0.9482
2	0.8912	0.6750	0.1674	0.5780	0.0517	0.8970
3	0.3826	0.9267	0.1667	0.7604	0.0371	0.9127
4	1.0000	0.9908	0.6417	0.9318	0.0332	0.9325

Table 3 Experimental results of data set 2

Method	CA_1	CA_2	CA_3	CA_4	CA_5	CA	EE	CC
1	1.0000	0.9944	0.9760	0.9615	0.9766	0.9860	0.0048	0.9904
2	0.8601	0.1389	0.9028	0.5455	0.9861	0.6866	0.0288	0.9484
3	0.9423	1.0000	0.9440	0.2115	0.7578	0.8855	0.0094	0.9812
4	1.0000	0.9944	0.9360	0.9231	0.9844	0.9777	0.0062	0.9877

Table 4 Experimental results of data set 3

Method	CA_1	CA_2	CA_3	CA	EE	CC
1	1.0000	1.0000	0.4783	0.9633	0.0176	0.9663
2	0.8073	0.8349	0.1376	0.5933	0.0319	0.9368
3	1.0000	0.3150	0.2174	0.6789	0.0274	0.9484
4	1.0000	0.9843	0.2609	0.9419	0.0216	0.9588

Although the proposed method can improve the accuracy of the classification of health state and the RUL prediction, the computational cost is inevitably increased because of the complexity of the method. The increased computational cost is caused by the clustering process of GMM and the feature selection of DET, which are generated prior to applying SVM. The time costs of GMM clustering process of data sets 1, 2, and 3 are 14.2389, 16.7022, and 15.3217 s, respectively. The time costs of feature selection of data sets 1, 2, and 3 are 7.1319, 7.7799, and 7.5327 s, respectively.

4 Conclusions

In this study, a new approach based on GMM and DET is proposed for predicting the RUL of rolling bearings. GMM is used to cluster the health states and identify the abnormal data sets from the training data sets. The MDL principle is used to determine the number of clusters for dividing all the data sets into several health states adaptively and practically. After obtaining the health states, salient features are selected by DET for increasing the classification and prediction accuracy. In the prediction process, LS-SVM is utilized to predict the RUL of rolling bearings by inputting the salient features. The experimental results indicate that the proposed method shows a reliable performance in predicting RUL and classifying the rolling bearings. This approach can also be applied to the prognosis of other mechanical assets. The future work will focus on seeking for improved and advanced feature extraction methods, clustering methods, and mode recognition methods to enhance the accuracy of RUL prediction. Considering that the complexity of the proposed method increases the computational cost, the computational efficiency will also be improved.

Acknowledgements The authors gratefully acknowledge the support of the National Key Research and Development Program of China (Grant No. 2016YFF0203400), the National Natural Science Foundation of China (Grant Nos. 51575168 and 51375152), the Project of National Science and Technology Supporting Plan (Grant No. 2015BAF32B03), and the Science Research Key Program of Educational Department of Hunan Province of China (Grant No. 16A180). The authors appreciate the support provided by the Collaborative Innovation Center of Intelligent New Energy Vehicle, the Hunan Collaborative Innovation Center for Green Car.

References

1. Marble S, Morton B P. Predicting the remaining life of propulsion system bearings. In: Proceedings of IEEE Aerospace Conference. IEEE, 2006, 1–8
2. Liao H, Zhao W, Guo H. Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model. In: Proceedings of IEEE Annual Reliability and Maintainability Symposium Conference. Newport Beach: IEEE, 2006, 127–132
3. Tian Z, Liao H. Condition based maintenance optimization for multi-component systems using proportional hazards model. Reliability Engineering & System Safety, 2011, 96(5): 581–589
4. Sikorska J Z, Hodkiewicz M, Ma L. Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing, 2011, 25(5): 1803–1836
5. Gebraeel N Z, Lawley M A, Liu R, et al. Residual life predictions from vibration-based degradation signals: A neural network approach. IEEE Transactions on Industrial Electronics, 2004, 51 (3): 694–700
6. Di Maio F, Tsui K L, Zio E. Combining relevance vector machines and exponential regression for bearing residual life estimation. Mechanical Systems and Signal Processing, 2012, 31(1): 405–427
7. Ben Ali J, Chebel-Morello B, Saidi L, et al. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. Mechanical Systems and Signal Processing, 2015, 56–57: 150–172
8. Pan D, Liu J, Cao J. Remaining useful life estimation using an inverse Gaussian degradation model. Neurocomputing, 2016, 185: 64–72
9. Zhao M, Tang B, Tan Q. Bearing remaining useful life estimation based on time-frequency representation and supervised dimensionality reduction. Measurement, 2016, 86: 41–55
10. Chen C, Vachtsevanos G, Orchard M E. Machine remaining useful life prediction: An integrated adaptive neuro-fuzzy and high-order particle filtering approach. Mechanical Systems and Signal Processing, 2012, 28: 597–607
11. Lu C, Chen J, Hong R, et al. Degradation trend estimation of slewing bearing based on LSSVM model. Mechanical Systems and Signal Processing, 2016, 76–77: 353–366
12. Loutas T H, Roulias D, Georgoulas G. Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression. IEEE Transactions on Reliability, 2013, 62(4): 821–832
13. Khanmohammadi S, Chou C A. A Gaussian mixture model based discretization algorithm for associative classification of medical data. Expert Systems with Applications, 2016, 58: 119–129

14. Elguebaly T, Bouguila N. Simultaneous high-dimensional clustering and feature selection using asymmetric Gaussian mixture models. Image and Vision Computing, 2015, 34: 27–41
15. Yu J. Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models. Mechanical Systems and Signal Processing, 2011, 25(7): 2573–2588
16. Heyns T, Heyns P S, de Villiers J P. Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox. Mechanical Systems and Signal Processing, 2012, 32: 200–215
17. Yang B S, Han T, Huang W W. Fault diagnosis of rotating machinery based on multi-class support vector machines. Journal of Mechanical Science and Technology, 2005, 19(3): 846–859
18. Zeng M, Yang Y, Zheng J, et al. Maximum margin classification based on flexible convex hulls for fault diagnosis of roller bearings. Mechanical Systems and Signal Processing, 2016, 66–67: 533–545
19. Lei Y, He Z, Zi Y, et al. New clustering algorithm-based fault diagnosis using compensation distance evaluation technique. Mechanical Systems and Signal Processing, 2008, 22(2): 419–435
20. Choi S W, Park J H, Lee I B. Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. Computers & Chemical Engineering, 2004, 28(8): 1377–1387
21. Lei Y, Lin J, He Z, et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing, 2013, 35(1–2): 108–126
22. Gai G. The processing of rotor startup signals based on empirical mode decomposition. Mechanical Systems and Signal Processing, 2006, 20(1): 222–235
23. Huang N E, Zheng S, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for non linear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1998, 454 (1971): 903–995
24. Yeh M H. The complex bidimensional empirical mode decomposition. Signal Processing, 2012, 92(2): 523–541
25. NASA. IMS bearings data set. 2014. Retrieved from http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/
26. Qiu H, Lee J, Lin J, et al. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. Advanced Engineering Informatics, 2003, 17(3–4): 127–140
27. Qiu H, Lee J, Lin J, et al. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. Journal of Sound and Vibration, 2006, 289(4–5): 1066–1090