IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCES IN PROGNOSTICS AND SYSTEM HEALTH MANAGEMENT

Received May 13, 2019, accepted June 25, 2019, date of publication July 5, 2019, date of current version July 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927104

Machinery Early Fault Detection Based on Dirichlet Process Mixture Model

BO MA[1,2], YI ZHAO[1,2], YING ZHANG[3], QING LEI JIANG[4,5], AND XIU QUN HOU[4,5]

[1]Beijing Key Laboratory for Health Monitoring and Self-Recovery of High End Mechanical Equipment, Beijing University of Chemical Technology, Beijing 100029, China
[2]Key Laboratory of Ministry of Education for Engine Health Monitoring and Networking, Beijing University of Chemical Technology, Beijing 100029, China
[3]School of Mechanical and Mechatronic Systems, University of Technology Sydney, Sydney, NSW 201101, Australia
[4]Research Institute of Nuclear Power Operations, Wuhan 430000, China
[5]China Nuclear Power Operation Technology Corporation, Ltd., Wuhan 430000, China

Corresponding author: Bo Ma (mabo@mail.buct.edu.cn)

This work was supported by the Generic Technology Research and Application of National Quality Infrastructure under Grant 2016YFF0203303.

ABSTRACT The most commonly used single feature-based anomaly detection method for the complex machinery, such as large wind power equipment, steam turbine generator sets, and reciprocating compressors, exhibits a defect of low-alarm accuracy due to the non-stationary characteristic of the vibration signals. In order to improve the accuracy of fault detection, a novel method based on the Dirichlet process mixture model (DPMM) is proposed. First, the features of the mechanical vibration signals are used to construct the feature space of the equipment. The DPMM modeling method is then applied to self-learn the probabilistic mixture model of the feature space. The normal working condition model is used as the benchmark model. The early fault detection is realized by using a precise difference measurement method based on Kullback–Leibler divergence to calculate the difference between the real-time model and the benchmark model accurately, and by comparing the calculation result with a self-learned alarm threshold. The effectiveness and the adaptability of this novel early fault detection method are verified by comparing it to the single feature-based anomaly detection method and the Gaussian mixture model (GMM)-based early fault detection method.

INDEX TERMS Early fault detection, Dirichlet process mixture model, machinery, vibration signal.

I. INTRODUCTION

Complex machinery such as large wind power equipment, steam turbine generator sets and reciprocating compressors are widely used in industrial production. Once the machinery breaks down, substantial economic losses occur. However, due to the non-stationary characteristics of the mechanical vibration signals, the most commonly used single feature-based anomaly detection method is characterized by low accuracy. Therefore, the improvement of the accuracy of fault detection for complex machinery has become a research hotspot in the field of condition monitoring.

In recent years, the artificial neural network (ANN), as an intelligent recognizer, has achieved good application results in the field of machinery fault detection [1], [2]. However, the insufficient fault samples give rise to the limited application of these methods. Support vector machine (SVM) and relevant vector machine (RVM) overcome the shortcomings of ANN, and have been extensively used in the field of machinery fault detection, which relies on their superiority in dealing with small samples [3]–[5]. However, the fault detection accuracy of these methods is greatly influenced by the kernel function. Without the complex optimal modeling of the kernel function, a high recognition rate is difficult to achieve. Li and Liang [6] realized gearbox multiple faults detection by using bi-component sparse low-rank matrix separation-based method. Sun *et al.* [7] proposed a wind turbine fault detection method based on multiwavelet denoising with the data-driven block threshold. Tian *et al.* [8] conducted motor bearing fault detection by using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis. Osman and Wang [9] proposed a fault detection method based on the morphological Hilbert-Huang transform technique. Because the response signals collected from the measuring point are

The associate editor coordinating the review of this manuscript and approving it for publication was Wen-Sheng Zhao.

the superposition and coupling of different excitation signals, the tiny changes of the excitation signals are difficult to reflect on the response signal. Although the aforementioned methods have improved the accuracy of the fault detection to a certain degree, they do not alarm at an appropriate time, and some even fail to discover some faults because they cannot identify the changes of the excitation signals. The change process of the excitation signal can be considered to follow an unknown distribution, so the distribution of the response signals is the linear superposition of each unknown distribution. If the distribution characteristics of the response signal can be characterized, the changes of the excitation signals will be accurately identified [10]. Li *et al.* [11] proposed a performance degradation assessment method based on Gaussian mixture model (GMM), that characterized the statistical distribution of the response signals by constructing the GMM of the signals. The bearing degradation process was accurately depicted via the analysis of the changes of the model. However, because the number of the components contained in the GMM must be set manually, the fitting effect of GMM on the response signals will be seriously affected if it is not set properly.

The Dirichlet process mixture model (DPMM) is a cornerstone of nonparametric Bayesian statistics. Due to the Dirichlet process provides a prior information for the distribution of the parameters of the mixture model, DPMM can accurately and automatically determine the number of the components contained in the model according to the observed data [12]. Because of this advantage, DPMM has been widely used in the data mining field for purposes such as text clustering and image segmentation [13]–[16] and has achieved excellent performance. Therefore, if DPMM can be applied to the statistical distribution self-learning of mechanical vibration signals, the accuracy of complex machinery early fault detection can be significantly improved .

In this paper, a novel early fault detection method based on DPMM is proposed for complex machinery. The method uses the features of mechanical vibration signals to construct the feature space of the equipment. The feature spaces under both the normal and real-time working conditions are further used as the training samples for DPMM to establish the benchmark model and the real-time model respectively. The early fault detection is realized by using a precise difference measurement method based on Kullback-Leibler (KL) divergence to accurately calculate the difference between the real-time model and the benchmark model, and by comparing the calculation result with a self-learned alarm threshold. The experimental results indicate the feasibility and effectiveness of the method.

The remainder of the paper is organized as follows. In Section II, the theory of the Dirichlet process mixture model is briefly explained. The proposed method is presented in Section III. In Section IV, the obtained results are provided and discussed. Finally, conclusions are drawn in Section V.

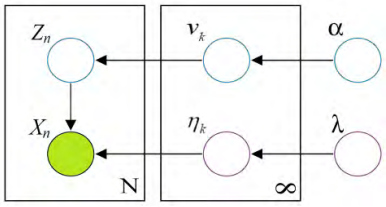

FIGURE 1. Graphical model of a DPMM.

II. DIRICHLET PROCESS MIXTURE MODEL

In this section, the theory of the Dirichlet process mixture model (DPMM) is presented in two steps. First, the DPMM is reviewed briefly. The parameters of the model are then elaborately derived by the variational Bayesian inference method.

A. REVIEW OF DIRICHLET PROCESS MIXTURE MODEL

If the parameters of a hierarchical model follow DP prior distribution, the model is called DPMM. DP is defined as the distribution of a set of distributions or random measures. It can be regarded as an extended infinite dimensional Dirichlet distribution. We consider two infinite collections of independent random variables $\boldsymbol{v} = \{v_k\}_{k=1}^{\infty}$ and $\boldsymbol{\eta} = \{\eta_k\}_{k=1}^{\infty}$, where $v_k \sim Beta(1, \alpha)$ and $\eta_k \sim G_0$. α and G_0 are the scale parameter and the basic distribution of DP, respectively. The stick-breaking representation of DP [17] is shown as follows:

$$\pi_k(\boldsymbol{v}) = v_k \prod_{l=1}^{k-1} (1 - v_l) \tag{1}$$

$$G = \sum_{k=1}^{\infty} \pi_k(\boldsymbol{v}) \delta_{\eta_k}. \tag{2}$$

where δ_{η_k} represents the distribution concentrated at a single point η_k.

The stick-breaking representation of DPMM is defined as

$$\boldsymbol{\eta} \,|G_0 \sim G_0 \tag{3}$$

$$\boldsymbol{v} \,|\alpha \sim Beta(1, \alpha) \tag{4}$$

$$Z_n \,|\boldsymbol{v} \sim \text{Mult}(\pi(\boldsymbol{v})) \tag{5}$$

$$X_n \,|z_n \sim p(X_n \,|\eta_{z_n}), \tag{6}$$

where Z_n is the indicating variable that follows the multinomial distribution, X_n is the observation data, $\boldsymbol{\pi}(\boldsymbol{v})$ represents the weight vector of the mixture model, and $\boldsymbol{\eta}$ denotes the parameter set of the mixture model which includes mean vector $\boldsymbol{\mu}$ and covariance matrix set $\sum$. $\pi(\boldsymbol{v})$ is used to generate the indicator vector Z_n, which can assign X_n to the specified distribution component.

The graphical model of a DPMM is illustrated in Fig. 1, in which nodes represent random variables and plates indicate the replication.

B. DERIVATION OF MODEL PARAMETERS

Consider a DPMM with hyperparameters $\boldsymbol{\theta} = \{\alpha, \lambda\}$ and latent variable $\boldsymbol{R} = \{\boldsymbol{v}, \boldsymbol{\eta}, \boldsymbol{Z}\}$. In order to obtain the posterior

distribution $p(\boldsymbol{R}|\boldsymbol{X},\boldsymbol{\theta})$, a distribution family $q_\varepsilon(\boldsymbol{R})$ indexed by the variational parameter ε is constructed. The posterior distribution can be approximated by using the mean field variation inference method. The target of model training is to minimize the KL divergence between $q_\varepsilon(\boldsymbol{R})$ and $p(\boldsymbol{R}|\boldsymbol{X},\boldsymbol{\theta})$, which can be expressed as follows:

$$D\,q_\varepsilon(\boldsymbol{R})\|p(\boldsymbol{R}|\boldsymbol{X},\boldsymbol{\theta})] = E_q\left[\log q_\varepsilon(\boldsymbol{R})\right] - E_q[\log p(\boldsymbol{R},\boldsymbol{X}|\boldsymbol{\theta})] + \log p(\boldsymbol{X}|\boldsymbol{\theta}). \quad (7)$$

Eq.(7) can be transformed as follows according to the non-negative characteristic of KL divergence is:

$$\log p(\boldsymbol{X}|\alpha,\lambda) \geq L, \quad (8)$$

with

$$L = E_q[\log p(\boldsymbol{v}|\alpha)] + E_q[\log p(\boldsymbol{\eta}|\lambda)] + \sum_{n=1}^{N}\left\{E_q\left[\log p\left(Z_n|\boldsymbol{v}\right)\right] + E_q\left[\log p\left(X_n|Z_n\right)\right]\right\} - E_q[\log q(\boldsymbol{v},\boldsymbol{\eta},\boldsymbol{Z})]. \quad (9)$$

The target of the model training can be transformed to maximize the variational low bound of the log marginal probability L.

By using the representation of the truncated stick-breaking process and the factorization hypothesis, the variational distribution family for the mean field variation inference can be defined as

$$q(\boldsymbol{v},\boldsymbol{\eta},\boldsymbol{Z}) = \prod_{t=1}^{T-1} q_{\gamma_t}(v_t)\prod_{t=1}^{T} q_{\tau_t}(\eta_t)\prod_{i=1}^{N} q_{\phi_i}(Z_n), \quad (10)$$

where $q_{\gamma_t}(v_t)$ denotes beta distributions, $q_{\tau_t}(\eta_t)$ represents exponential family distributions, and $q_{\phi_i}(z_n)$ indicates multinomial distributions. The truncated level T is a variational parameter that can be set freely. According to Eq.(10), we have

$$\varepsilon = \{\gamma_1,\ldots,\gamma_{T-1},\tau_1,\ldots,\tau_T,\phi_1,\ldots,\phi_N\}. \quad (11)$$

Repeatedly updating the variational parameter ε will increase the low bound L and finally acquire a local maximum.

III. PROPOSED EARLY FAULT DETECTION METHOD

The response function of the mechanical response point [18] is shown as follows:

$$Y(t) = \sum_{j=1}^{J} F_j(t)H_j(t), \quad (12)$$

where $F_j(t)$ represents the j^{th} function of the excitation force at time t, and $H_j(t)$ denotes the transfer function of the transfer path, which is between the j^{th} excitation source and the response point. The change process of the mechanical excitation signal obeys an unknown distribution and therefore the distribution of the response signal is a linear superposition of each unknown distribution. If a fault occurs in the equipment, the distributions of some of the excitation signals will change due to the changes of the excitation forces or the transfer paths, and the distribution of the response signals will vary accordingly. The DPMM can accurately characterize the distribution of the mechanical response signals, which rely on its ability to automatically determine the number of components. Therefore, the change of the mechanical running status can be recognized by analyzing the change of the DPMM.

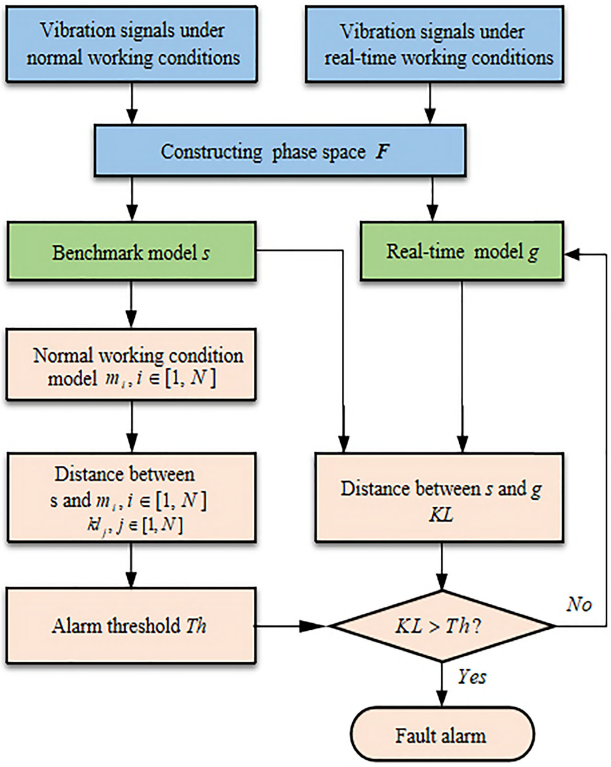

FIGURE 2. Flowchart of the early fault detection method.

The flowchart of the novel early fault detection method is illustrated in Fig. 2. This method is primarily comprised of three parts: feature space construction, training of the model, and self-learning of the alarm threshold.

A. FEATURE SPACE CONSTRUCTION

The feature space $\boldsymbol{F} = \{\boldsymbol{F_a}\}_{a=1}^{b}$ is constructed by the features of mechanical vibration signals, where b denotes the number of the datasets contained in each training sample, and $\boldsymbol{F_a}$ represents the feature matrix of the a^{th} dataset, which can be defined as

$$\boldsymbol{F}_a = \begin{bmatrix} f_{1,1} & \cdots & f_{1,p} \\ \vdots & \ddots & \vdots \\ f_{q,1} & \cdots & f_{q,p} \end{bmatrix}_{q\times p}, \quad (13)$$

where $f_{i,j}$ is the j^{th} feature of the signals collected from the i^{th} measuring point, q is the number of measuring points, and p is the number of features. Because the feature space contains a lot of information on the running status of the equipment, if it is used as the training sample for DPMM, the tiny changes of the running status can be noticeably reflected on the model.

B. TRAINING OF THE MODEL

The feature space $\boldsymbol{F}$ is used as the training sample of the DPMM modeling method, which corresponds to the observed data set $\boldsymbol{X}$ in Eq.(6). The posterior distribution $p(\boldsymbol{R}\,|\boldsymbol{X},\boldsymbol{\theta})$ of the hidden variable $\boldsymbol{R}=\{\boldsymbol{v},\boldsymbol{\eta},\boldsymbol{Z}\}$ is then approximated via the mean field variation inference method. The parameters of the model can be acquired when the training target of the model is achieved. The DPMM under the normal working conditions is used as the benchmark model, which can be defined as follows according to Eqs.(3-6):

$$\boldsymbol{\eta}_s\,|G_0 \sim G_0 \tag{14}$$

$$\boldsymbol{v}_s\,|\alpha \sim Beta(1,\alpha) \tag{15}$$

$$Z_{s,n}\,|\boldsymbol{v}_s \sim \mathrm{Mult}(\pi(\boldsymbol{v}_s)) \tag{16}$$

$$\boldsymbol{F}_{s,n}\,|z_{s,n} \sim p(\boldsymbol{F}_{s,n}\,|\eta_{s,z_{s,n}}), \tag{17}$$

where $\boldsymbol{F}_s$ denotes the feature space under normal working conditions, and $\boldsymbol{\eta}_s$ and $\boldsymbol{\pi}(\boldsymbol{v}_s)$ denote the parameter vector and the weight vector of the benchmark model, respectively.

C. SELF-LEARNING OF THE ALARM THRESHOLD

The proposed method conducts the early fault detection by calculating the difference between the real-time model g and the benchmark model s, and comparing the calculation result with a self-learned alarm threshold. Therefore, determining how to accurately measure the difference between s and g is the key to improving the accuracy of the early fault detection. The matching-based KL divergence approximation method is a representation method of the distance (i.e., the difference) between two probabilistic mixture models [19]. This method can gradually approximate the real difference between the two models by measuring the KL divergence (i.e. the difference) between the corresponding components of the two models one by one. If the proposed method uses the matching-based KL divergence approximation method to measure the distance between s and g, it can effectively reflect the tiny changes of the model, and thereby improve the effect of the early fault detection. The distance between s and g is shown as follows,

$$KL(s\,\|g)=\sum_{i=1}^{N}\pi(\boldsymbol{v}_s)_i[KL(s_i\,\|g_{w(i)})+\log\frac{\pi(\boldsymbol{v}_s)_i}{\pi(\boldsymbol{v}_g)_{w(i)}}] \tag{18}$$

with

$$w(i)=\arg\min_j[KL(s_i\,\|g_j)-\log\eta_{g,j}], \tag{19}$$

where N is the number of the components contained in s, and $\pi(\boldsymbol{v}_s)_i$ and $\pi(\boldsymbol{v}_g)_{w(i)}$ denote the weights of the i^{th} component of s and the $w(i)^{th}$ component of g, respectively. $KL(s_i\,\|g_j)$ is given by

$$KL\left(s_i\|g_j\right)=\frac{1}{2}\left[\log\frac{|\Sigma_j|}{|\Sigma_i|}+\mathrm{Tr}\left(\Sigma_j^{-1}\Sigma_i\right)+\left(\mu_i-\mu_j\right)^T\Sigma_j^{-1}\left(\mu_i-\mu_j\right)\right], \tag{20}$$

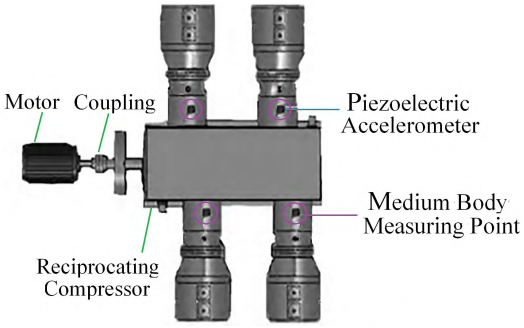

FIGURE 3. Layout of the sensors.

where μ_i and Σ_i denote the mean vector and the covariance matrix of the i^{th} component of s, respectively, μ_j and Σ_j represent the mean vector and the covariance matrix of the j^{th} component of g, respectively, and $Tr(\Sigma_j^{-1}\Sigma_i)$ indicates the trace of $\Sigma_j^{-1}\Sigma_i$.

When the running status of the equipment is the same, the distances between different DPMMs fluctuate slightly. Therefore, we can assume that these distances approximately obey a normal distribution. According to this characteristic, we can determine the alarm threshold by using the 3σ criterion. Here we assume that $m_i, i \in [1,N]$ represents N *DPMMs* under the normal working conditions, and $kl_j, j \in [1,N]$ denotes the distances between s and $m_i, i \in [1,N]$. Because the probability of the values reaching out of the range of $(\mu-3\sigma,\mu+3\sigma)$ in a normal distribution is less than 0.3%, the alarm threshold Th can be set to $\mu_{kl}+3\sigma_{kl}\mu_{kl}+3\sigma_{kl}$ after calculating the average value μ_{kl} and standard deviation σ_{kl} of $kl_j, j \in [1,N]$. When the distance between g and s exceeds Th, the method considers that a fault occurs in the equipment, and then further triggers an alarm.

IV. EXPERIMENTAL EVALUATION

In order to evaluate the performances of the proposed method, an experiment on early fault detection was conducted on reciprocating compressors. It is designed to detect 30 sets of actual fault case data, which contain three fault types.

The vibration signals of the reciprocating compressors are collected by the piezoelectric accelerometers. The layout of the sensors is illustrated in Fig. 3. The rotary speed of the motor is 300rpm, and the reciprocating compressor is D-type with 4-cylinders (4-D). The piezoelectric accelerometers are installed on the 4 medium body measuring points of the reciprocating compressor. The sampling frequency of the piezoelectric accelerometer is 10240Hz and the sampling length is two operation cycles of the reciprocating compressor (i.e. 0.4s).

All fault case data were collected from the production sites, and each described the entire degradation process of the reciprocating compressor from a normal working condition to an anomaly working condition. The details of the fault case data are listed in Table 2. To make a tradeoff between the

TABLE 1. Description of features.

Domain	Types of features	The number of features
Time domain	Peak value; RMS value; Kurtosis index; Impulsion index; Tolerance index	5
Time-frequency domain	Peak values of the eight frequency bands obtained by three-layer wavelet packet decomposition; RMS values of the eight frequency bands obtained by three-layer wavelet packet decomposition	16
Angle domain	The distances between 180° and the angles which correspond to the peak values of the data, which is within the range of 30°centered on the valve opening and closing angles; RMS values of the data, which is within the range of 30° centered on the valve opening and closing angles; The distance between 180°and the corresponding angle of the peak value; The distances between 180°and the angles which correspond to the peak values of the data, which is within the range of 30° centered on the commutation angles; RMS values of the data, which is within the range of 30° centered on the commutation angles	13

TABLE 2. Description of fault case data.

Types of faults	Types of reciprocating compressors	The number of fault cases
Piston assembly wear	2-M	11
	2-D	5
Valve leakage	6-M	2
	4-M	3
	3-D	1
	2-D	5
Liquid strike	2-L	1
	2-D	2

effect of the early fault detection and the training efficiency of the model, the sample size was set to 400 when dividing the fault case data. The first normal sample and the first 20 normal samples in each set of fault case data were selected to train the benchmark model and to self-learn the alarm threshold, respectively. The remaining samples were used for testing.

When the original vibration signals are used to train the DPMM, DPMM can most accurately characterize the running status of the reciprocating compressor, which makes the proposed method work best. However, in order to improve the efficiency of calculation, features are extracted from vibration signals as comprehensively as possible. In this experiment, 21 common features for mechanical anomaly detection were extracted from time domain and time-frequency domain [20], [21], and extra 13 angle domain features were added according to the motion characteristics of the reciprocating compressors. The types of the features are provided in Table 1. All of the 34 features of the vibration signals were used to construct the feature space.

In the experiment, the GMM-based early fault detection method (GMM method) and the single feature-based anomaly detection method (SF method) were compare with the proposed method (DPMM method). The number of the components contained in GMM was set to 3, which is a commonly used setting that can achieve better experimental results. The detection results of the three methods are shown in Table 3, from which it can be seen that the early fault detection accuracy of the DPMM method attains 93.33%, which is the highest among the three methods. Although the GMM method can conduct early fault detection, the accuracy of this method is low. The SF method is obviously unable to alarm at the early stage of the faults. In addition, the calculating efficiencies of both the DPMM and GMM methods are compared in Table 3. The time consumptions of the DPMM and GMM methods for training a model are 2.31s and 2.65s, respectively. The DPMM method requires less time consumption, and therefore it has a higher efficiency. The average lengths of time for early warning of both the DPMM and GMM methods are further compared in Table 4. It can be observed that the DPMM method alarms ahead of the GMM method for all sets of test data. The details of the detection results of both methods are exhibited in Table 5. It can be seen that the GMM method can accurately detect the valve leakage fault when the type of the reciprocating compressors is 2-D, and the detection accuracy of the same fault is low for other types of equipment. However, the DPMM method can accurately conduct the detection of valve leakage for all types of the equipment. Therefore, the DPMM method has stronger adaptability.

The detection effects of both the DPMM and GMM methods for the above three types of faults are shown in Figs. 4-6. The dotted lines in the figures indicate the self-learned alarm thresholds, and the circles denote the positions where the distance between the benchmark model and the real-time model exceeds the alarm threshold. It can be clearly observed that the DPMM method

TABLE 3. Detection results.

Methods	The number of test cases	The number of cases that can be detected early	Early detection accuracy (%)	Time consumption during model training (s)
DPMM	30	28	93.33	2.31
GMM	30	22	73.33	2.65
SF	30	0	0	—

TABLE 4. Average length of time for early warning.

Methods	Average length of time for piston assembly wear (h)	Average length of time for valve leakage (h)	Average length of time for liquid strike (h)
DPMM	85.25	16.20	6.62
GMM	50.30	12.54	3.62

TABLE 5. Details of detection results of DPMM and GMM methods.

Types of faults	Types of reciprocating compressors	The number of test cases	The number of cases that can be Detected early	
			DPMM	GMM
Piston assembly wear	2-M	11	9	7
	2-D	5	5	4
Valve leakage	6-M	2	2	1
	4-M	3	3	2
	3-D	1	1	0
	2-D	5	5	5
Liquid strike	2-L	1	1	1
	2-D	2	2	2

TABLE 6. Effect of training sample size on the average length of time for early warning.

Training sample size	Average length of time for piston assembly wear (h)	Average length of time for valve leakage (h)	Average length of time for liquid strike (h)
400	85.25	16.20	6.62
500	86.10	16.45	6.65
600	86.54	16.60	6.67

alarms before the GMM method for the three types of faults.

All the above comparative analyses verify the superiority of the DPMM method in early fault detection as compared to the GMM and SF methods. Because DPMM can self-learn the number of components from the data, it can more accurately represent the distribution of the feature space than can GMM. Therefore, the DPMM method more easily identifies the changes of the excitation signals than does the GMM method.

To explore the effect of training sample size on early fault detection, the DPMM method was tested 12 times.

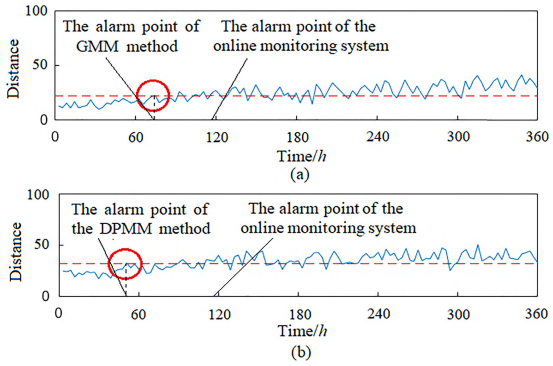

FIGURE 4. Detection effect of piston wear assembly. (a) GMM method. (b) DPMM method.

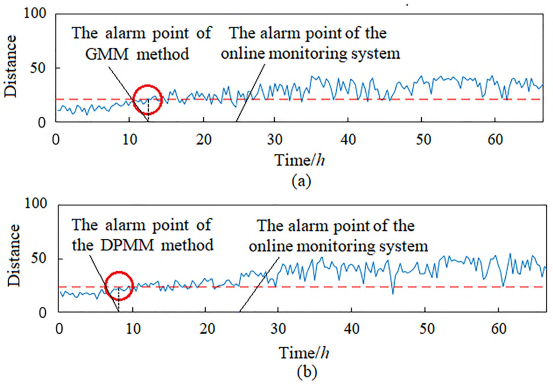

FIGURE 5. Detection effect of valve leakage. (a) GMM method. (b) DPMM method.

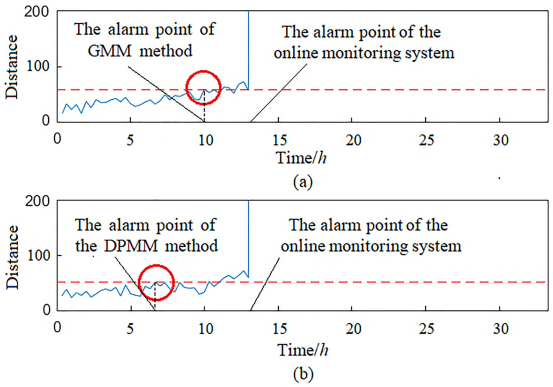

FIGURE 6. Detection effect of liquid strike. (a) GMM method. (b) DPMM method.

The training sample size in each test ranged from 50 to 600 with an interval of 50. The test results are provided in Fig. 7, from which it is evident that the training sample size is linear with the time consumption during model training. However, the accuracy of the DPMM method increases slowly when the sample size exceeds 400. Therefore, by comprehensively considering the accuracy and the efficiency of the method, the reasonable sample size is set to 400. The average lengths of time for early warning when the training sample size is equal to or exceeds 400 are further compared in Table 6. It can be seen

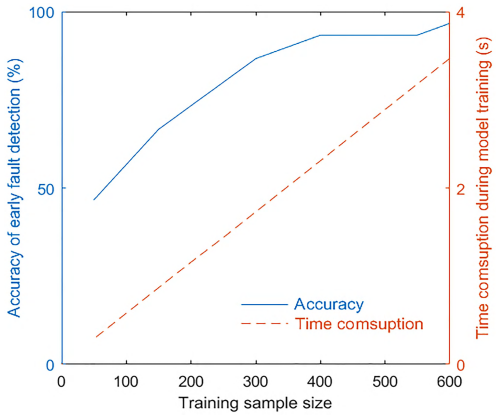

FIGURE 7. Effect of training sample size on the effect of early fault detection.

that this index is less affected by the training sample size.

V. CONCLUSIONS

In this paper, a novel early fault detection method based on the Dirichlet process mixture mode (DPMM) is proposed for complex machinery. Because most of the traditional fault detection methods are unable to analyze the changes of the excitation signals, they are unable to alarm at an appropriate time or even fail to discover some faults. In this work, the distribution of the mechanical response signals is characterized by DPMM, which can improve the fitting effect on the response signals by self-learning the number of the components contained in the model. The feature space is constructed and used as the training samples for the DPMM modeling method to increase the sensitivity to the changes of the running status of the DPMM. The early fault detection is realized by using a precise difference measurement method based on Kullback-Leibler (KL) divergence to accurately calculate the difference between the real-time model and the benchmark model, and by comparing the calculation results with a self-learned alarm threshold.

The actual fault case data of the reciprocating compressors is used to verify the effectiveness of the proposed method. The results confirm that the proposed method can accurately early detect the typical faults of complex machinery and can greatly advance the alarm time point.

However, it should be noted that the early fault detection effect of the proposed method is difficult to achieve the best due to the features extracted by the traditional feature extraction methods are difficult to contain rich information of the original vibration signals. Feature work will use the deep learning algorithm to extract features adaptively.

REFERENCES

[1] H. Shao, H. Jiang, X. Li, and T. Liang, "Rolling bearing fault detection using continuous deep belief network with locally linear embedding," *Comput. Ind.*, vol. 96, pp. 27–39, Apr. 2018.

[2] A. Zabihi-Hesari, S. Ansari-Rad, F. A. Shirazi, and M. Ayati, "Fault detection and diagnosis of a 12-cylinder trainset diesel engine based on vibration signature analysis and neural network," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 233, no. 6, pp. 1910–1923, 2019. doi: 10.1177/0954406218778313.

[3] J. A. Carino, M. Delgado-Prieto, J. A. Iglesias, A. Sanchis, D. Zurita, M. Millan, J. A. O. Redondo, and R. Romero-Troncoso, "Fault detection and identification methodology under an incremental learning framework applied to industrial machinery," *IEEE Access*, vol. 6, pp. 49755–49766, 2018.

[4] L. Selak, P. Butala, and A. Sluga, "Condition monitoring and fault diagnostics for hydropower plants," *Comput. Ind.*, vol. 65, no. 6, pp. 924–936, Aug. 2014.

[5] W. J. An, M. G. Liang, and H. Liu, "An improved one-class support vector machine classifier for outlier detection," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 229, pp. 580–588, Feb. 2015.

[6] Q. Li and S. Y. Liang, "Multiple faults detection for rotating machinery based on bicomponent sparse low-rank matrix separation approach," *IEEE Access*, vol. 6, pp. 20242–20254, 2018.

[7] H. Sun, Y. Zi, and Z. He, "Wind turbine fault detection using multiwavelet denoising with the data-driven block threshold," *Appl. Acoust.*, vol. 77, pp. 122–129, Mar. 2014.

[8] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, Mar. 2016.

[9] S. Osman and W. Wang, "A morphological Hilbert–Huang transform technique for bearing fault detection," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 11, pp. 2646–2656, Nov. 2016.

[10] B. Ma, Y. Zhao, and Y. Zhang, "Application of variational auto-encoder in mechanical fault early warning," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chongqing)*, Oct. 2018, pp. 1263–1268.

[11] W. H. Li, B. X. Dai, and S. H. Zhang, "Bearing performance degradation assessment based on Wavelet packet entropy and Gaussian mixture model," *J. Vib. Shock*, vol. 32, no. 21, pp. 35–40, 2013.

[12] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *Ann. Statist.*, vol. 1, no. 2, pp. 209–230, Mar. 1973.

[13] R. Z. Huang, G. Yu, Z. J. Wang, J. Zhang, and L. Shi, "Dirichlet process mixture model for document clustering with feature partition," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 8, pp. 1748–1759, Aug. 2013.

[14] R. Granell, C. J. Axon, and D. C. H. Wallom, "Clustering disaggregated load profiles using a Dirichlet process mixture model," *Energy Convers. Manage.*, vol. 92, no. 4, pp. 507–516, Mar. 2015.

[15] W. Song, M. Li, P. Zhang, Y. Wu, L. Jia, and L. An, "Unsupervised PolSAR image classification and segmentation using Dirichlet process mixture model and Markov random fields with similarity measure," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 10, no. 8, pp. 3556–3568, Aug. 2017.

[16] B. J. Reich and H. D. Bondell, "A spatial Dirichlet process mixture model for clustering population genetics data," *Biometrics*, vol. 67, no. 2, pp. 381–390, Jun. 2011.

[17] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Anal.*, vol. 1, no. 1, pp. 121–143, 2006.

[18] X. J. Wu, Y. D. Lv, and F. S. Sui, "Basic theory of operational transfer path analysis and its application," *Noise Vibra. Contr.*, vol. 34, no. 1, pp. 28–31, Feb. 2014.

[19] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of KL-Divergence between two Gaussian mixtures," in *Proc. 9th IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2003, pp. 487–493.

[20] A. Moosavian, G. Najafi, B. Ghobadian, and S. M. Mirsalim, "The effect of piston scratching fault on the vibration behavior of an IC engine," *Appl. Acoust.*, vol. 126, pp. 91–100, Nov. 2017.

[21] L. Saidi, J. B. Ali, E. Bechhoefer, and M. Benbouzid, "Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR," *Appl. Acoust.*, vol. 120, pp. 1–8, May 2017.

BO MA received the Ph.D. degree in mechanical engineering from the Beijing University of Chemical Technology, Beijing, in 2006, where he is currently an Associate Professor with the Diagnosis and Self-Recovery Engineering Research Center. He is a distinguished scholar in the field of machinery condition monitoring and fault diagnosis. From 2013 to 2014, he was a Visiting Scholar with the University of California, Berkeley, USA. He has been an inventor of many patents. He has coauthored lots of technical papers published in international journals and conferences. His technical interests include expert systems, machine learning, and signal processing.

YI ZHAO received the B.S. degree in mechanical engineering from the Beijing University of Chemical Technology, Beijing, China, in 2016, where he is currently pursuing the M.S. degree in mechanical engineering. His technical interests include fault detection and diagnosis for reciprocating machinery, and probabilistic mixture model.

YING ZHANG received the B.S. degree in mechanical engineering and the M.S. degree in mechanical engineering from the Beijing University of Chemical Technology, Beijing, China, in 2014 and 2017, respectively. She is currently pursuing the Ph.D. degree with the School of Mechanical and Mechatronic Systems, University of Technology Sydney, NSW, Australia. Her technical interests include equipment condition monitoring and fault diagnosis, and deep learning.

QING LEI JIANG received the Ph.D. degree in mechanical engineering from Zhejiang University, Zhejiang, China, in 2012. He is currently with China Nuclear Power Operation Technology Corporation, Ltd., and the Research Institute of Nuclear Power Operation. His technical interests include machinery fault diagnosis and vibration control.

XIU QUN HOU received the M.S. degree in nuclear power generation engineering from Wuhan University, Wuhan, in 2017. He is currently with China Nuclear Power Operation Technology Corporation, Ltd., and the Research Institute of Nuclear Power Operation. His technical interest includes machinery intelligent fault diagnosis.