IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 24, 2018, accepted January 11, 2019, date of publication January 18, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2893277

Rolling Bearing Performance Degradation Assessment Based on Convolutional Sparse Combination Learning

WENYI HUANG, JUNSHENG CHENG, AND YU YANG
State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China

Corresponding author: Junsheng Cheng (chengjunsheng@hnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51575168 and Grant 51875183, in part by the Key Research and Development Program of Hunan Province under Grant 2017GK2182, in part by the Collaborative Innovation Center of Intelligent New Energy Vehicle, and in part by the Hunan Collaborative Innovation Center for Green Car.

ABSTRACT A novel bearing performance degradation assessment method based on convolutional sparse coding and combination learning is proposed in this paper, which can avoid the impact of the traditional features on the assessment results under complex operation conditions. The vibration signal can be decomposed into the convolution of the kernel sets and their corresponding sparse solutions using convolutional sparse coding. The learned kernel sets based on the training signal samples are the comprehensive embodiment of the information related to the operational complexity and the bearing healthy state, and the corresponding sparse solutions indicate the energy of the kernel activation. A simulation experiment of bearing signal proves that the activation energy of the kernels, which are more related to bearing healthy, will rise with the increase of the bearing degradation degree; meanwhile, the error of reconstructed signal based on the learned kernel sets and original signal will decrease since the description effect of convolutional sparse coding to the signal will be enhanced with the periodic strengthen of signal caused by the fault. Thus, an index based on the kernel sparse norms and the errors between the original and reconstructed signals has been proposed to evaluate the degradation degree of the bearing in this paper. On the other hand, combination learning will be fused into the convolutional sparse coding to improve the real-time performance of the index calculation and obtain the best assessment result in the testing part by dividing the kernel dictionary set into multiple sub-dictionaries. A bearing run-to-fail experiment is analyzed to verify the validity of the proposed method. The testing results show that the proposed assessment index can clearly detect the initial fault, serve fault and failure of the bearing in the whole life, and the proposed method is more self-adaptive and sensitive to the fault degree. In addition, it is verified that the time consumption of the combination dictionaries is less than that of a single dictionary set.

INDEX TERMS Bearing performance degradation assessment, complex operation condition, convolutional sparse combination learning, convolutional sparse reconstruction.

I. INTRODUCTION

With the rapid development of science and technology and the continuous development of industrial applications, mechanical equipment is becoming more and more complex, precise and intelligent. Therefore, the requirements of industrial condition monitoring and fault diagnosis system are higher and higher in practical application and many researchers dedicated to the study of engineering machinery degradation model. For example, Wang *et al.* [1] etc provided statistics of published papers related to multivariate statistical process monitoring (MSPM) over the past decade because MSPM methods such as principal component analysis (PCA), partial least squares (PLS) and independent component analysis (ICA) are significant for improving production efficiency and enhancing safety. In addition, they [2] also proposed a novel control performance assessment (CPA) method for iterative learning control-controlled batch processes based on a 2-D linear quadratic Gaussian (LQG) benchmark. The methods mentioned above have contributed to the theoretical research of maintenance and monitoring of engineering machinery.

No matter in heavy machinery or light machinery, the rolling bearing is the critical element of rotating machinery since its fault probability is 30% in all faults of rotating elements [3]. It is significant to monitor the bearing healthy on the working state of rotating machine. Accordingly, many researchers pay more and more attention to bearing fault detection [4], remaining useful life prediction [5] and performance degradation assessment in recent years. Especially in bearing performance degradation assessment, lots of work about this topic such as novel performance degradation indexes, feature optimization methods and feature fusion methods had been done, which provided potential directions for future research [6]–[10].

Undoubtedly, the feature extraction of bearing vibration signal is the key to accurately evaluate the bearing state. The common features include the time domain features such as mean, kurtosis, skew-ness and root mean square [11]–[13], frequency domain features such as mean frequency, center frequency [14]–[16] and time-frequency domain features such as wavelet packet energy, wavelet packet entropy and component feature of empirical mode decomposition [17]–[22]. Their effectiveness directly affects the availability of diagnosis model. However, the time domain features and frequency domain features are easily affected by noise, and the time-frequency domain features based on wavelet transform also depends on the prior knowledge of wavelet selection, which cannot make the bearing fault diagnosis methods based on traditional features applicable to any working environment.

With the rise of big data and the rapid development of artificial intelligence, compressive sensing (CS) has been widely used and has obtained state-of-the-art results in multiple fields such as machine learning, neuroscience, signal processing, image and audio processing, classification, and statistics [23]–[25] because its features are flexibility, sparse and super-resolution. CS mainly consists two parts including sparse coding and dictionary design. Sparse coding transforms signal into a linear combination of basis kernels in a redundant dictionary. It had been proved that the sparse coding is a NP-hard problem [26] and it can be replaced by the pursuit-based approximate solutions. The pursuit algorithms mainly include greedy-based matching pursuit and convex optimal-based basis pursuit. It has been proved that the convex optimal-based basis pursuits are more suitable for processing large-scale and high-dimensional engineering signals [27]–[30]. Dictionary design is, as much as possible, making the kernels learned from the data well match the impulses embedded in the signal. The general methods for constructing a dictionary include using a manually predefined dictionary via signal transformation methods, and adaptively learning a dictionary from the measured data itself (also called explicit dictionary). The predefined dictionaries, such as frequency dictionaries, time-scale dictionaries and time-frequency dictionaries are widely used for computational conveniences. However, an explicit dictionary is directly inferred from the input data by machine learning techniques, which can adapt to various kinds of operation conditions without prior knowledge. Shift-invariant dictionary learning (SIDL) is an extension of dictionary learning and it allows each kernel to be shifted at each time offset within the signal, which is suitable for extracting the impulses induced by the rotational machine faults [31]. For rotating mechanical fault diagnosis, CS based pursuit algorithms and SIDL can extract the circular impulses submerged in the vibration signals of a rotating machine system, which contain high-level structure features. The extracted impulses have highly adaptability to the complex conditions and can depict the detail information of the signal, which are superior to the traditional features. In addition, SIDL also has successfully been used to extract double impulses embedded in the bearing vibration signals. Thus, SIDL is a powerful signal processing method for coding impulse signals [32].

In the framework of CS, a novel convolutional sparse coding (CSC) based on alternating-direction method of multiplier (ADMM) was proposed in 2016 [33]. Not only does it have the excellent characteristics of SIDL but also can enhance the ability of the shift-invariant sparse coding (SISC). The ADMM based CSC had been applied in image processing [34]. Meanwhile, CSC has the clear physical meaning of representing the impulses generated by the circular interactions of faults. Therefore, CSC was applied to fault detection of wheel set bearing in a high-speed train by Jianming Ding. The decoding results of the bearing vibration signals showed that the proposed method can not only effectively detect the wheel set bearing fault but also accurately characterize the fault dynamic behaviors of the bearing defects [35]. But the application of CSC to performance degradation assessment of machine elements has not been proposed yet. Affirmatively, the CSC can be used to evaluate the fault degree because the activation times and strength of the learned impulses with more relationship to fault will rise as the fault deepens. Furthermore, with the periodic enhancement of the signal caused by the fault, the description effect of the CSC will be strengthened and the errors between the original and reconstructed signals will be decreased. Thus, in this paper, a bearing performance degradation method based on CSC has been proposed. The assessment index combines the impulses activation norm (sparse norms) and signal reconstructed error.

However, there are two disadvantages of CSC based methods. One is the long time consumption of sparse representation in the training and testing parts, which decreases the efficiency of on-line monitoring. Another one is the single dictionary representation of signal, which decreases the diversity of learned features and will lead the degradation assessment to be local optimum. Thus, in order to deal with the two issues, in this paper, combination learning is fused into the CSC (CSCL) to deduce the calculation complexity of the sparse model by dividing the impulse kernel dictionary set into multiple sub-dictionaries. This combination can not only reduce the time consumption of the computation in real-time, but also provide more assessment indexes for selection within the several sub-dictionaries. A bearing run-to-fail test will be used to verify the validity the performance degradation assessment method based on CSCL.

II. BACKGROUNDS

A. CONVOLUTIONAL SPARSE CODING

Models based on sparse coding are widely used in many fields such as image processing [36] and fault diagnosis [37]. Since the rolling bearing is periodic rotation, the features hidden in the vibration signal are periodic. Periodic features are usually considered to be the same, and they can be extracted by convolution sparse coding (CSC) because CSC has the characteristic of translation invariance. The CSC can be stated as a bi-convex problem as follows

$$\begin{aligned} \min_{d,x} \ & \frac{1}{2}\sum_{n=1}^{N}\left\| y_n - \sum_{k=1}^{R} d_k \otimes x_{k,n}\right\|_2^2 \\ & + \beta \sum_{n=1,k=1}^{N,R} \left\|x_{k,n}\right\|_p \\ s.t. \ & \|dk\|_2^2 \le 1 \end{aligned} \tag{1}$$

where y_n is the observed sample, $\{y_1, \ldots y_n, \ldots, y_N\}$ is the training set, $x_{k,n}$ are sparse coding representations, d_k are the corresponding convolution kernels, R is the number of the kernels, $\{x_{1,n}, \ldots x_{k,n}, \ldots x_{R,n}\}$ are the sparse representations of y_n based on the kernels$\{d1, \ldots d_k, \ldots d_R\}$, $\otimes$ is the convolution operator, β is the sparse weight and it is set to be 1 and p is set to be 2 in this paper because the L_2 norm stands for the energy of the sparse solution, which is beneficial for analyzing the energy distribution of each kernel. Analogously to standard sparse coding and other machine learning approaches, the convolution kernels can be learned from the training data by solving the optimization problem Eq. (1) as stated above. The role of kernel learning is to estimate a priori signal distribution. Then in the feature extraction phase, the kernels are fixed and the features are computed only by minimization over feature maps, which correspond to Bayesian inference from noisy measurements.

In sparse coding, the efficient methods of kernel learning had been presented in many literatures [38], [39] by alternately minimize over the feature maps while keeping the filters fixed and over the filters while keeping the feature maps fixed, taking the advantage that both sub-problems are convex, which can make the optimal problems convergence. In this way the feature extraction is essentially run to convergent in the iterations of the kernel learning algorithm.

B. ADMM

ADMM is suitable for solving the distributed convex optimization problems. The large global problem can be decomposed into multiple smaller local sub-problems which are easily solved by decomposing and coordinating, and then the large global solution can be obtained through the solution of the coordination sub-problems. The optimization problem can be stated as

$$\begin{aligned} & \min f(x) + g(z) \\ & s.t.\ Ax + Bz = c \end{aligned} \tag{2}$$

where $x \in \boldsymbol{R}^s$, $z \in \boldsymbol{R}^n$, $A \in \boldsymbol{R}^{p\times s}$, $B \in \boldsymbol{R}^{p\times n}$, $c \in \boldsymbol{R}^p$, $f(\cdot)$ and $g(\cdot)$ are convex functions, x and z are arguments, A and B are coefficients and c is a constant. The iterative process of solving the Eq. (2) by using ADMM can be stated as follows

$$\begin{aligned} x^{k+1} &= \arg\min f(x) + \frac{\rho}{2}\left\|Ax + Bz^k - c + u^k\right\|_2^2 \\ z^{k+1} &= \arg\min g(z) + \frac{\rho}{2}\left\|Ax^{k+1} + Bz - c + u^k\right\|_2^2 \\ u^{k+1} &= u^k + (Ax^{k+1} + Bz^{k+1} - c) \end{aligned} \tag{3}$$

where k is the iteration times of ADMM, x^k is the value of x in the k^{th} iteration, z^k is the value of z in the k^{th} iteration, u is the dual variable, and ρ is the penalty factor. In the iterations of ADMM, the values of z^k and u^k are calculated in the $k-1^{\text{th}}$ iteration. Firstly, the x^{k+1} is updated in the first equation of Eq. (3). Then the z^{k+1} is calculated based on the x^{k+1} in the second equation of Eq. (3). Finally, the u^{k+1} is updated based on the third equation of Eq. (3).

C. COMBINATION LEARNING

If only one dictionary is generated in large-scale training set, the number of kernels in the dictionary will become huge, which will make the process of searching for kernels in large dictionary time-consuming. In order to solve the problem of low search efficiency, combinatorial learning has been used to search for kernels in this paper. The concept of combinatorial learning [40] is to generate a dictionary set consisting of multiple compact sub-dictionaries from large-scale training data, and then search for the most matching sub-dictionary in parallel in the testing data. It can effectively reduce the time complexity of the model by convert a huge dictionary into multiple compact sub-dictionaries. On the other hand, for bearing performance degradation assessment, the divided sub-dictionaries will generate different assessment results since the testing sample has different sparse representations using the sub-dictionaries. Thus, more selections can be provided for the best performance assessment result using this diversity expression of signal.

D. ADAPTIVE THRESHOLD

The bearing performance degradation assessment index is a continuously variable parameter. It indicates the degree of bearing performance deviating from the normal condition. Setting the alarm threshold of the index is beneficial for monitoring operation state of the bearing. According to the 3σ rule in probability and statistics, the probability of a Gauss random variable with a mean value of x and a variance of σ^2 falling in the interval $(x - 3\sigma, x + 3\sigma)$ is 99.73%. Once a value exceeds this range, it is reasonable to consider that the value does not belong to the original scope. It is assumes that the degradation indexes with the similar degradation degree conform to a normal distribution. When the indexes exceed the interval constructed by the previous data, it can be confirmed that the degradation degree of the bearing has been changed greatly. Since the degradation index is monotonically increasing, the upper limiter of the threshold is set to be $x + 3\sigma$ and the adaptive segmental threshold is shown in Eq. 4.

$$Th(t) = \begin{cases} t = 1, \ldots, t_{ts}: \\ mean(T(1:t_{ts}) + 3std(T(1:t_{ts})), \\ t = t_{ts}+1, \ldots, t_e: \\ mean(T(1:t-1) + 3std(T(1:t-1)) \\ t = t_e+1, \ldots end: \\ mean(T(1:t_e-1) + 3std(T(1:t_e-1)) \end{cases} \tag{4}$$

where $T(t)$ is the degradation index of t^{th} sample, and t_{ts} is the training sample of initial threshold. The solution of Th is divided into three segments. The Th in the first segment is a fixed value, which comes from the normal condition. In the second segment, the $T(t)$ is compared with the $Th(t-1)$. If the $T(t)$ still remains in the normal stage, then it will be added into the normal data. However, if N_u consecutive $T(t)$s after $t = t_e$ exceed the upper limiter, then t_e is considered to be the time of initial fault.

III. BEARING PERFORMANCE DEGRADATION ASSESSMENT BASED ON CSCL

The bearing performance degradation assessment method based on CSCL is presented as follows and it includes learning part and testing part. In the learning part, the healthy bearing samples are used as the input of CSCL to generate several kernel sub-dictionaries which contain the detail information of the operation condition and healthy condition. In the testing part, the later bearing samples of the whole life are sparse decoded using the kernels of each sub-dictionary. The sparse representations and the errors of reconstructed signal are used to construct the assessment index of bearing degradation.

A. LEARNING PART

The goal of convolution sparse combinatorial learning is to generate a dictionary set S, where $S = \{s_1, \ldots, s_m, \ldots s_M\}$, s_m is the sub-dictionary, M is the iterative times and the number of sub-dictionary. The sub-dictionary s_m is generated to cover as many as possible training samples in the m^{th} iteration, and then it is used for updating the kernel and sparse representation.

Assuming that the current state is the m^{th} iteration, the following three parts are the processing of generating sub-dictionary: (1) the solution of sparse representation, (2) kernel learning, (3) updating of training set and dictionary set $\boldsymbol{S}$.

1) THE SOLUTION OF SPARSE REPRESENTATION

The dictionary of the equation (1) is replaced by a sub-dictionary s_m

$$\min_{x_n} \frac{1}{2}\sum_{n=1}^{\overline{N}} \|y_n - S_m x_n\|_2^2 + \beta \sum_{n=1}^{\overline{N}} \|x_n\|_p \tag{5}$$

where S_m is the combination of sub-dictionary s_m and convolution operator $\otimes$, $x_n = [x_{1,n}, \ldots, x_{1,n}, \ldots x_{L,n}]$ is the sparse representation of sample y_n, L is the kernel number in sub-dictionary s_m, and $\overline{N}$ is the sample number of the m^{th} iteration. The Eq. (5) can be split by using ADMM into:

$$\begin{aligned} \sum_{n=1}^{\overline{N}} f(x_n) &= \frac{1}{2}\sum_{n=1}^{\overline{N}} \|y_n - S_m x_n\|_2^2 \\ \sum_{n=1}^{\overline{N}} g(z_n) &= \beta \sum_{n=1}^{\overline{N}} \|z_n\|_p \\ s.t. x_n - z_n &= \vec{0}, n \in [1, \overline{N}] \end{aligned} \tag{6}$$

Then Eq. (6) will be added into Eq. (3)

$$x_n^{k+1} = \arg\min \sum_{n=1}^{\overline{N}} (\frac{1}{2}\|y_n - s_m x_n\|_2^2 + \frac{\rho}{2}\left\|x_n - z_n^k + u_n^k\right\|_2^2) \tag{7}$$

$$z_n^{k+1} = \arg\min \sum_{n=1}^{\overline{N}} (\beta\|z_n\|_p + \frac{\rho}{2}\left\|x_n^{k+1} - z_n + u_n^k\right\|_2^2) \tag{8}$$

$$u_n^{k+1} = u_n^k + x_n^{k+1} - z_n^{k+1} \tag{9}$$

where Eq. (7) is a quadratic form problem and the derivative of optimal solution is zero.

$$\begin{aligned} x_n^{k+1} = (I/\rho - (1/\rho)^2 S_m^T \\ \cdot (S_m S_m^T/\rho + I)^{-1} S_m)(S_m^T y_n + \rho a) \end{aligned} \tag{10}$$

where $a = z_n^k - u_n^k$, I is the identity matrix, ρ is the penalty factor.

Eq. (8) is the lp-norm minimization problem of non-convex [41], which can be solved by using generalization of soft-threshold

$$z_n^{k+1} = \begin{cases} 0, & |a| \le \tau(\varphi) \\ sgn(a) \cdot T(|a|;\varphi), & |a| > \tau(\varphi) \end{cases} \tag{11}$$

$$\begin{aligned} \tau(\varphi) = (2\varphi(1-p))^{1/(2-p)} \\ + \varphi p(2\varphi(1-p))^{(p-1)/(2-p)}) \end{aligned} \tag{12}$$

$$T(|a|;\varphi) - |a| + \varphi p(T(|a|;\varphi))^{p-1} = 0 \tag{13}$$

where $a = x_n^{k+1} + u_n^k$, $\psi = \beta/\rho$, sgn(·) is the symbolic function.

The ADMM iterations of Eq. (7), Eq. (8) and Eq. (9) are the same as the process of Eq. (3) in part 2.2. The algorithm for optimization over sparse representation, which also serves as the feature extraction algorithm is summarized in Algorithm 1.

2) KERNEL LEARNING

The x_n will be fixed in this part

$$\begin{aligned} \min_{s_m} \sum_{n=1}^{\overline{N}} \frac{1}{2} \|y_n - X_n \cdot s_m\|_2^2 \\ s.t.\ sm = [s_{1,m}, \ldots, s_{L,m}]^T, \quad \|sl,m\|_2^2 \le 1 \end{aligned} \tag{14}$$

Algorithm 1 Sparse Representations Solving

1:Initialize: $k = 0$, $z_n^0 = 0$, $u_n^0 = 0$;
2:Pre-compute Fourier transforms $|\hat{s}_{l,m}|$;
3:**while**
4:Update x_n^{k+1}: using formula (10);
5: Update z_n^{k+1}: using formula (11);
6:Update u_n^{k+1}: using formula (9);
7: **until convergence**
8:Output: Sparse representation$\{x_n\}$, $n \in [1, \bar{N}]$.

Algorithm 2 Kernel Learning

1:Initialize: $k = 0$, $u_m^0 = 0$, $v_m^0 = s_m^0$, s_m^0 is randomly;
2:**while**
3:Update $\{x_n\}$: using Algorithm 1;
4:**while**
5:Update $x_{n,m}^{k+1}$: using formula (18);
6: Update v_m^{k+1}: using formula (19);
7:Update u^{k+1}: using formula (17);
8: **until convergence**
9:until convergence
10:Output: Sparse representation $\{x_n\}$ and dictionary s_m.

Then the Eq. (14) can be split using ADMM

$$\sum_{n=1}^{\bar{N}} f(s_m) = \sum_{n=1}^{\bar{N}} \frac{1}{2} \|y_n - X_n \cdot s_m\|_2^2$$
$$g(v_{l,m}) = \begin{cases} 0, & v_{l,m} \in C \\ +\infty, & \text{otherwise} \end{cases}$$
$$s.t.\ s_m - v_m = \vec{0} \tag{15}$$

where s_m and v_m are sub-dictionary, $v_{1,m}$ is the kernel of v_m, $v_m = [v_{1,m}, \ldots, v_{L,m}]^T$, $\boldsymbol{C}$ is the convex set, $\boldsymbol{S}$ is the support set of $v_{1,m}$. Then Eq. (15) is added into Eq. (3).

$$s_m^{k+1} = \arg\min_{s_m} \sum_{n=1}^{\bar{N}} (\frac{1}{2} \|y_n - X_n s_m\|_2^2 + \frac{\rho}{2} \left\| s_m - v_m^k + u_m^k \right\|_2^2) \tag{16}$$

$$v_{l,m}^{k+1} = \arg\min_{v_{l,m}} (g(v_{l,m}) + \frac{\rho}{2} \sum_{n=1}^{\bar{N}} \left\| s_{l,m}^{k+1} - v_{l,m} + u_{l,m}^k \right\|_2^2) \tag{17}$$

$$u_m^{k+1} = u_m^k + s_m^{k+1} - v_m^{k+1} \tag{18}$$

where $s_m = [s_{1,m}, \ldots, s_{L,m}]^T$, $v_m = [v_{1,m}, \ldots v_{L,m}]^T$, and $u_m = [u_{1,m}, \ldots, u_{L,m}]^T$.

The solution of Eq. (16) is the same with Eq. (7):

$$s_m^{k+1} = (I/\rho - (1/\rho)^2 X_n^T \cdot (X_n X_n^T/\rho + I)^{-1} X_n)(X_n^T y_n + \rho a) \tag{19}$$

The Eq. (17) can be solved by using approximate operator

$$v_{l,m}^{k+1} = \prod_C (b) = \begin{cases} b/\|b\| & \text{if } \|b \cdot T\| > 1 \\ b \cdot T & \text{otherwise} \end{cases} \tag{20}$$

$$b = \frac{1}{\bar{N}} \sum_{n=1}^{\bar{N}} s_{l,m}^{k+1} + v_{l,m}^k \tag{21}$$

$$T = \begin{cases} 1, & \textit{if } \text{supp(b)} \in \text{s} \\ 0, & \textit{otherwise} \end{cases} \tag{22}$$

where the $\prod_c$ (b) represents the mapping of b to the convex set C.

The algorithm for optimization over kernels is summarized in Algorithm 2.

3) UPDATING OF TRAINING SET AND DICTIONARY SET

The sparse representation $\{x_n\}$ and sub-dictionary s_m in the m^{th} iteration can be obtained after part (A) and part (B), and then the convolution sparse reconstructed signal can be calculated using x_n and s_m. Since the kernels are learned from the training sets of healthy bearing signals in the whole life of bearing, the degradation assessment index is closely related to convolution sparse reconstruction cost C. Suppose that the kernel number of sub-dictionary s_m is L, the cost C of the training sample y_n based on sub-dictionary s_m can be defined as:

$$C_{n,m} = \frac{1}{2} \left\| y_n - \sum_{l=1}^{L} s_{l,m} \otimes x_{l,n} \right\|_2^2 + \beta \sum_{l=1}^{L} \frac{1}{L} w_{l,m} \left\| x_{l,n} \right\|_p \tag{23}$$

$$w_k = 1 - \frac{t_k}{\sum_{k=1}^{R} t_k} \quad tk = \sum_{n=1}^{N} \left\| x_{k,n} \right\|_2^2 \tag{24}$$

where $w_{1,m}$ is the weight of each kernel. The main basis of the weight design is that the kernel of lower activation energy is more related to the degradation in the healthy training samples. Thus the bigger the t_k is, the smaller the w_k is.

The updating of training set and dictionary $\boldsymbol{S}$ can be summarized in Algorithm 3, where

$$\lambda_m = 1.05^m \cdot \lambda_0 \tag{25}$$
$$\lambda_0 = R(c_{n,m=1}, \gamma)$$
$$s.t.\ \gamma \in [10\%, 50\%] \tag{26}$$

R is the ascending symbol, and λ_0 is equal to the γ^{th} element in the ascending order of $c_{n,m-1}$.

B. TESTING PART

The dictionary set S and the weight set W are obtained in the learning part. Then, they will be used to construct a performance degradation assessment index in the testing part.

The solution of sparse representation in testing part is similar to part (A) in the learning part. The sparse representation x_t of testing sample y_t based on the sub-dictionary s_m

Algorithm 3 The Update of Training Set and Dictionary Set

1:Initialize: $m = 1$, $Y_{m=1}$ is the training set of initial iteration, dictionary assemble $S = \varphi$, weight assemble $W = \varphi$, λ_m is the threshold;
2:**while**
3:Obtain $\{x_n\}$, s_m: using Algorithm 2;
4: update S: adding s_m into S;
5:Caculate $c_{n,m}$, w_m: using formula (23),(24);
6: Update W: adding w_m into W;
7:Update dictionary: $m++$;
8: Update training
set:: $\mathrm{Y_m} = \begin{cases} \text{add } \mathrm{y_n} \text{ into } \mathrm{Y_m}, \text{if } \mathrm{C_{n,m}} > \lambda_\mathrm{m} \\ \text{No operation, otherwise} \end{cases}$
9:untilY_m decrease to φ
10:Output: Dictionary assemble S, weight assemble W.

can be stated as

$$\min_{x_t} \frac{1}{2} \|y_t - S_m x_t\|_2^2 + \beta \|x_t\|_p \tag{27}$$

$$x_t^{k+1} = \left(I/\rho - (1/\rho)^2 s_m^T (s_m s_m^T/\rho + I)^{-1} s_m\right)(s_m^T y_t + \rho a_x) \tag{28}$$

$$z_t^{k+1} = \begin{cases} 0, & \text{if } |a_z| \le \tau(\varphi) \\ sgn(a_z) \cdot T(|a_z|; \varphi), & \text{otherwise} \end{cases} \tag{29}$$

where $a_x = z_t^k - u_t^k$, $a_z = x_t^{k+1} + u_t^k$ and I is the identity matrix.

With the fault degree increasing, the activation energy $||x_t||^{\mathrm{p}}$ of kernels related to fault degree must be rising. However, the activation energy of the kernels unrelated to fault may decrease or keep stable. On the other hand, with the periodic enhancement of the signal caused by the fault, the description effect of the CSC is strengthened and the reconstructed signal must be more similar to the original signal. To verify the correctness of the above views, a simulation experiment of the bearing with different fault degree will be presented. The simulation signal of bearing outer race fault $s(t)$ is simply designed as Eq. (30) [42], [43], where y_0 is the displacement constant, ξ is the damping coefficient, f_n is the nature frequency, f_r is the rotation frequency, f_d is the bearing fault frequency, α is the coefficient between f_r and f_d since there is a linear relationship between the bearing fault frequency and rotation frequency, f_s is the sampling frequency, N is the number of sampling points, $n(t)$ is Gaussian white noise signal obeying standard uniform distribution on the open interval $(0, m^2)$, and λ is the amplitude coefficient of the noise amplitude.

$$y(t) = \sum_{m=1}^{K} y_0 \cdot e^{-2\pi\xi f_n (t - \frac{m}{f_d}) \sin[2\pi\xi f_n (t - \frac{m}{f_d})]} \tag{30}$$

$$K = ceil(\frac{N \cdot f_d}{fs}) \tag{31}$$

$$f_d = \alpha \cdot f_r \tag{32}$$

$$n(t) \sim N(0, m^2) \tag{33}$$

$$s(t) = y(t) + \lambda \cdot n(t) \tag{34}$$

Set $\xi = 0.1, f_n = 1e3, f_s = 10e3, f_r = 30, \alpha = 2, \lambda = 0.5$, $N = 2000$, the simulation signal of bearing with different fault degree can be obtained by changing y_0. In this section, bearing signals of four fault degree ($y_0 = 0.4$; $y_0 = 0.8$; $y_0 = 1.2$; $y_0 = 1.6$) are simulated and shown in Fig. 1. There are 50 sets of signal samples in each fault degree.

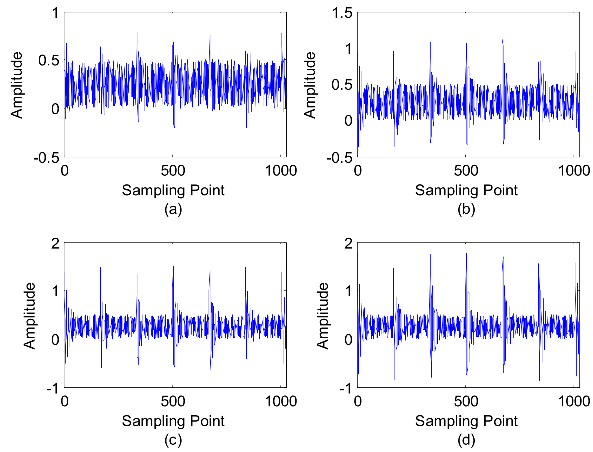

FIGURE 1. Simulation signal of different fault degree. (a) y0 = 0.4; (b) y0 = 0.8; (c) y0 = 1.2; (d) y0 = 1.6).

The signal samples shown in Fig. 1 (a) with $y_0 = 0.4$ are used for kernel learning using the Algorithm 1∼3. The scales of the each kernel are: four kernels of 80 length, four kernels of 120 length, and four kernels of 160 lengths. In order to verify that the kernels learned from the bearing signals of early stage can be used for feature extraction and signal reconstruction when the fault degree goes deeper. The active energy and reconstructed errors of the bearing signal with the other three fault degree (b), (c), (d) are calculated based on the learned kernels. The average active energies of the twelve kernels are shown in Fig. 2 and the reconstructed signals are shown in Fig. 3. The specific values of the active energy and the reconstructed errors (REs) are shown in Table 1.

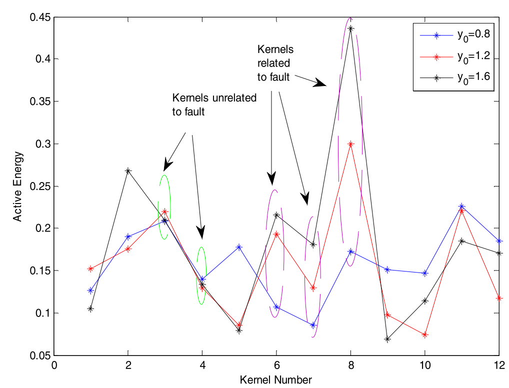

FIGURE 2. Kernel active energy of the three fault degree.

It can be inferred from the Fig. 2 that the active energies of the 6th, 7th and 8th kernel keep increasing with the deepening of fault degree. However, the active energies of the third

TABLE 1. Active energy and the reconstructed errors.

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	**RE**
(b)	0.12	0.19	0.20	0.14	0.17	0.10	0.08	0.17	0.15	0.14	0.22	0.18	**24.57**
(c)	0.15	0.17	0.21	0.12	0.08	0.19	0.12	0.30	0.09	0.07	0.22	0.11	**23.02**
(d)	0.10	0.26	0.20	0.13	0.08	0.21	0.18	0.43	0.06	0.11	0.18	0.17	**21.47**

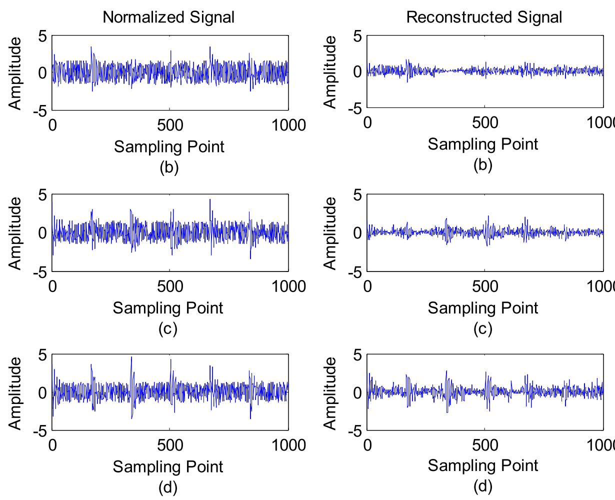

FIGURE 3. Normalized signals and reconstructed signals of different fault degree. (b) y0 = 0.8; (c) y0 = 1.2; (d) y0 = 1.6.

and fourth kernel basically remain the same. Thus, the 6th, 7th, 8th kernel can be considered to be related to the fault degree. On the other hand, it can be seen from the Fig. 2 that the reconstructed signals with deeper fault degree are more similar to the original signals and the REs keep decreasing with the increasing of fault degree in the Table 1.

Based on the analyses of CSC to the bearing signal of different fault degrees above, the cost $C_{t,m}$ of testing sample y_t based on sub-dictionary can be defined as

$$C_{t,m} = -N(\frac{1}{2}\left\| y_t - \sum_{l=1}^{L} s_{l,m} \otimes x_{l,t} \right\|) + \beta \cdot N(\sum_{l=1}^{L} \overline{w}_{l,m,t} \cdot w_{l,m} \cdot \left\| x_{l,t} \right\|_p) \quad (35)$$

where N is the normalized operator, $\bar{w}_{\mathrm{l,m,t}}$ is the degradation tendency weight, which can be defined as in Eq. (36).

$$\overline{w}_{l,m,t} = \begin{cases} 1/L, & if\ t \leq L\_train \\ \dfrac{CF(N(T_{l,m,t}),2) \cdot entropy(N(T_{l,m,t}))}{\sum_{l=1}^{L} CF(N(T_{l,m,t}),2) \cdot entropy(N(T_{l,m,t}))}, & otherwise \end{cases}$$

$$T_{l,m,t} = \left[\left\| x_{l,m,1} \right\|_p, \left\| x_{l,m,2} \right\|_p, \cdots, \left\| x_{l,m,t} \right\|_p \right] \quad (36)$$

where entropy(·) is the Shannon entropy which is used to control the stability of the cost, CF(·, 2) is the quadratic term coefficient of quadratic function fitting. Since the norm tendency of different kernels is unpredictable, $\bar{w}_{\mathrm{l,m,t}}$ is designed to make the changing tendency of each kernel norm consistent.

For the same sample, the sub-dictionary of higher cost is considered to be more sensitive to the fault degree than other sub-dictionaries because the fault degree is monotonically increasing over the time in theory. Choosing the higher cost as the assessment index can improve the incrementing of the index and detect the early failure earlier. Thus, the performance degradation assessment index φ of testing sample y_t can be defined as

$$\phi = \max\{C_{t,1}, \ldots, C_{t,m}, \ldots C_{t,M}\} \quad (37)$$

The whole performance degradation assessment process of rolling bearing can be summarized in Algorithm 4 and the specific process is shown in Fig. 4.

Algorithm 4 Bearing Performance Degradation Assessment

1:Input: Dictionary assemble S, Weight assemble W, testing set y_T;
2:for $t = 1 \ldots T$
3:for $m = 1 \ldots M$
4: while
5: Update x_t^{k+1}: using formula (25);;
6: Update z_t^{k+1}: using formula (26);
7: Update u_t^{k+1}: $u_t^{k+1} = u_t^k + x_t^{k+1}$-$z_t^{k+1}$;
8:Update $\bar{w}_{\mathrm{l,m,t}}$: using formula (29);
9:Obtain the cost $C_{t,m}$: using formula (27);
10:until end
11:Obtain the performance degradation assessment index φ of y_t: using formula (30);
12:**until** end

IV. EXPERIMENTAL STUDY

To verify the effectiveness of the proposed performance degradation assessment method based on CSCL, a bearing run-to-fail experiment is analyzed [44]. The experiment performed bearing run-to failure tests under constant load conditions on a specially designed test rig as shown in Fig. 5. The bearing test rig hosts four test Rexnord ZA-2115 double row bearing on one shaft. The shaft was driven by an AC motor and coupled by rub belts. The rotation speed was kept constant at 2000 rpm. A radial load of 6000 lbs was added to the shaft and the bearing by a spring mechanism.

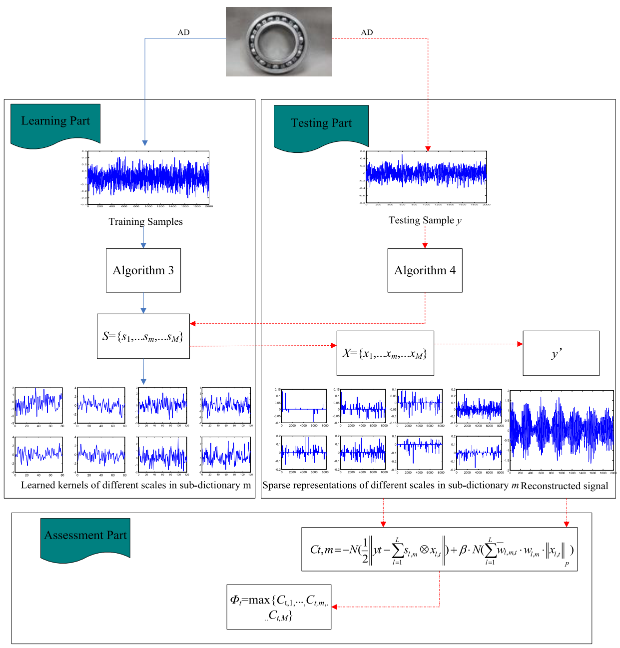

FIGURE 4. **Algorithm flow of bearing performance degradation assessment based on CSCL.**

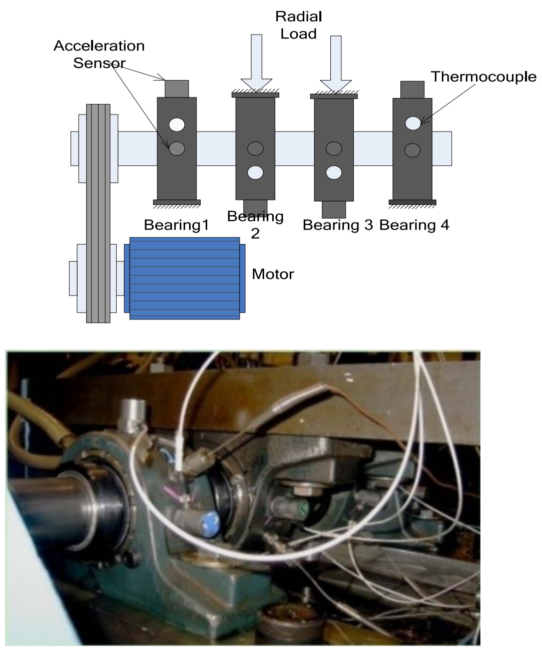

FIGURE 5. **Schematic and photo of test rig.**

Data collection started form the 2004.2.12 10:32:39 to 2004.2.19 06:22:39 with no interruption in the acquisition process. Collecting the vibration signals every 10 minutes during acquisition time, 984 data files were altogether collected during the experimental process. The sampling frequency is 20 KHz and each sensor collects 20480 sampling points each time. This paper analysis the first 8192 sampling points of the second file (Bearing 2) and only one vibration sensor was applied for signal collection.

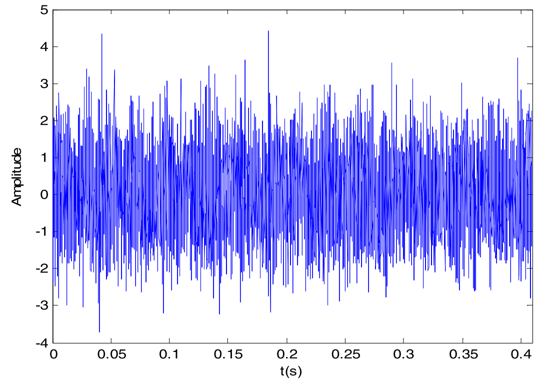

FIGURE 6. **Time-plot of the 100th training sample.**

A. ESTABLISHMENT AND VERIFICATION OF TRAINING MODEL

Before evaluating the degradation degree of bearing, the training model is established based on the healthy bearing samples. In this experiment, the training set is the first 200 samples. The dictionary set can be obtained after the training set is learned using CSCL. To verify the validity of the model, the sparsity of the sparse representations and the time-frequency comparisons of the original normalized signal and reconstructed signal are analyzed, respectively.

1) SPARSITY OF THE SPARSE REPRESENTATION

Two dictionaries are generated in the training set, the kernels of three scales in the first dictionary $s1$ are shown in Fig. 7 (a), (c) and (e) (The second dictionary is not shown for the reason of space).The scales of the kernels are the same as the testing part. The 80, 120 and 160 sampling points are corresponding to the bearing rotation of 0.133, 0.167 and 0.267 times which means that the kernels are transient. The original bearing signal of the 100th sample is shown in Fig. 6 and it must be normalized using zero-mean normalization before training. The sparse representations of signal based on the kernels are shown in Fig. 7 (b), (d) and (f). It can be seen from the Fig. 7 that the some "thorns" occurred in the sparse representation and the activation times of different kernels are variable, which can be concluded that the activation of the kernels is sparse and the sub-dictionary is valid.

2) TIME-FREQUENCY COMPARISONS

The normalized original signal and reconstructed signal based on the learned kernels are shown in Fig. 8 (a). The rotation frequency (2000rpm/min) can be found from the reconstructed signal since the adjacent pulse is 0.03s. On the other hand, the time-frequency magnitude scalograms of the original and reconstructed signal are shown in Fig.8 (b). The time-frequency analysis method is wavelet transform. The type of the wavelet is Morlet, whose bandwidth parameter and center frequency are both 3. It can be seen from the Fig. 8 (b) that the color distributions of the two magnitude scalograms are almost the same, which means that the

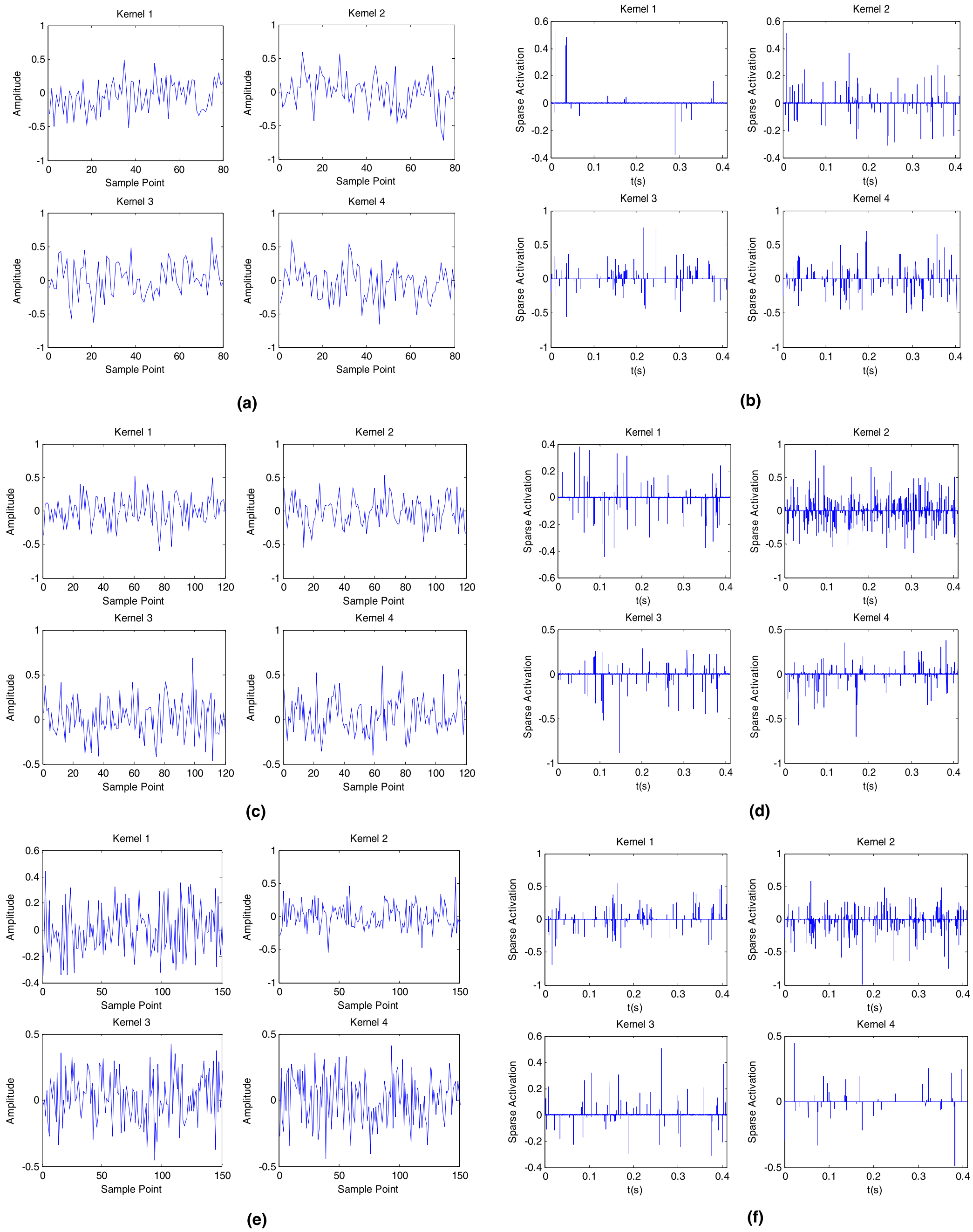

FIGURE 7. Kernels of s1 and the corresponding sparse representations of the100th sample. (a) Kernels of scale I. (b) Sparse representations of kernels of scale I. (c) Kernels of scale II. (d) Sparse representations of kernels of scale II. (e) Kernels of scale III. (f) Sparse representations of kernels of scale III.

normalized original signal and the reconstructed signal have a high similarity in the time- frequency domain. It is also verified that the sub-dictionary used for reconfiguration is effective.

B. PERFORMANCE DEGRADATION ASSESSMENT OF ROLLING BEARING

After trainging the assessment model, the testing samples of the whole bearing life are used to evaluate the bearing degradation degree by using Algorithm 4. The sparse norms of the twelve kernels are shown in Fig. 9(a). It can be seen from the figure that the variation trends of the sparse norms are different with the increasing of bearing fault in the sub-dictionary. Some of the trends keep stable, some increase and the others decrease. It can be concluded that the kernels have different sensitive to the fault degree. Thus, the trend weight of each kernel must be time-varying to make all the kernel norms have a unified trend. The weights of the kernels

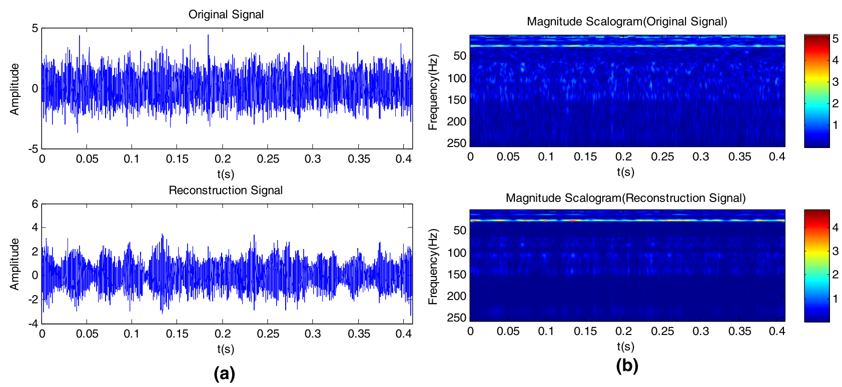

FIGURE 8. Comparisons of original signal and reconstructed signal. (a) Time domain. (b) Time-frequency domain.

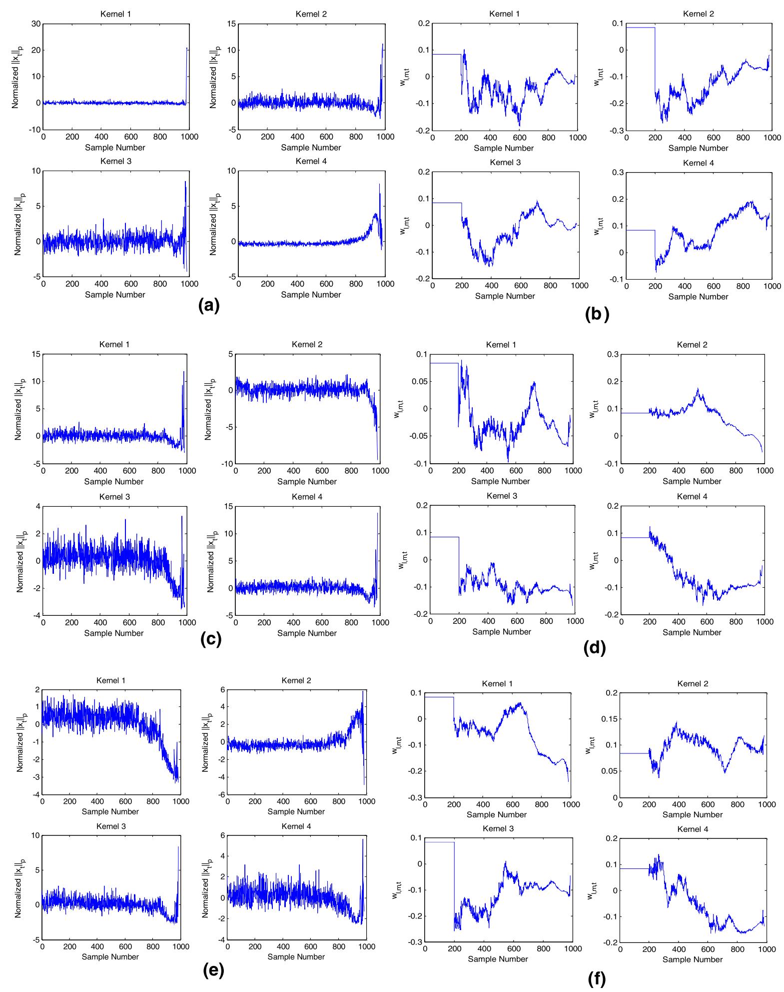

FIGURE 9. Sparse norm tendencies and the corresponding time-varying weights of the kernels in s1. (a) Norm trends of kernels of scales I. (b) Time-varying weights of kernels of scales I. (c) Norm trends of kernels of scales II. (d) Time-varying weights of kernels of scales II. (e) Norm trends of kernels of scales III. (f) Time-varying weights of kernels of scales III.

in each scale are shown in Fig. 9(b). The weights are time-varing with the change of the sparse norm. The fusion sparse index and error index of $s1$ are shown in Fig. 10. The two indicators are smoothly and have clear trends in the whole bearing life. Finally, the performance degradation assessment index is fused by the two indicators of the two sub-dictionary

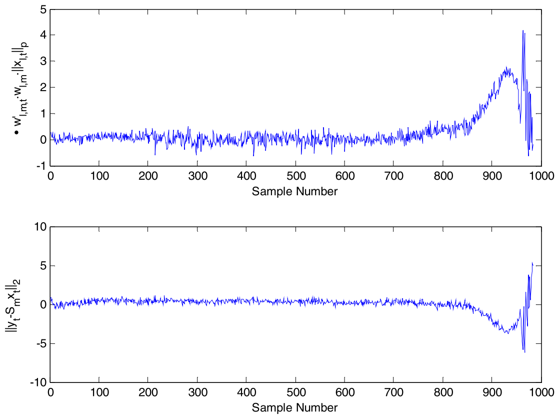

FIGURE 10. Two normalized performance degradation indicators of s1 in the whole life of bearing.

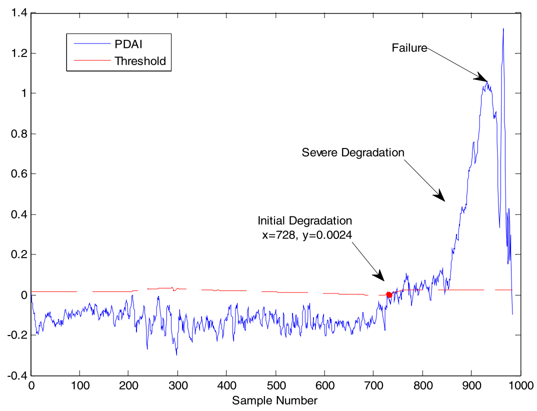

FIGURE 11. The fusion performance degradation index φ of dictionary set S.

and it is shown in Fig.11. The paremeter of the threshold is set to be: $t_{ts} = 200, N_u = 10$. It can be seen from the Fig. 11 that: (1) The assessment index c increases with the time goes on in general, which is consistent with the trend of bearing degradation; (2) The initial degradation can be detected when $t = 728$; (3) The degradation is deepening with the φ rising continuously and failure of the bearing occurrs when the φ has a sharp fluctuation. The results above prove that the proposed index is effective and the training model is reasonable, which can provide a feasible assessment programme for the bearing fault degree detection in practice.

C. COMPARISON

To verify the advantages of the self-adaptive features extracted by the proposed method, two performance degradation assessment methods based on traditional feature extraction and clustering model are presented. Twenty-three traditional features used for bearing fault diagnosis are extracted from the first 200 samples of the bearing signal to construct the healthy clustering model. These features contain time-domain features, frequency-domain features and time-frequency domain features [21], [22]. In this paper, the Gaussian Mixed Model (GMM) [45] and the Support vector data description (SVDD) [46] are used as clustering models. The Euclidean distance between the sphere centre of SVDD model and the features of the testing sample is considered as the distance degradation index. Also, the normalized negative log likelihood probability of the testing sample features subordinated to the GMM model is considered as the probability degradation index. The bearing performance degradation assessment indexes based on the two methods above are respectively shown in Fig. 12 and Fig. 13. Compared with the proposed method, the performance indexes based on traditional feature process methods are not smooth enough and hard to detect the initial fault because the features based on the priori knowledge are easily affected by the working conditions or environment noise. However, the proposed method based on CSCL can adaptively extract the information related to the fault, which makes the degradation index more stable and more sensitive to the fault degree.

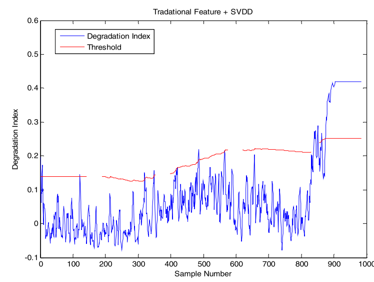

FIGURE 12. Degradation assessment index based on traditional features and SVDD.

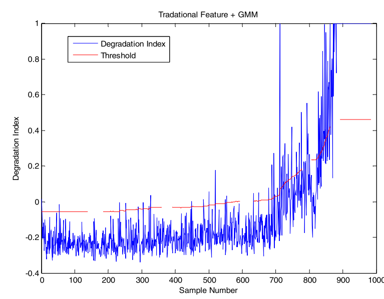

FIGURE 13. Degradation assessment index based on traditional features and GMM.

D. TIME-CONSUMING COMPARISONS OF SINGLE DICTIONARY AND COMBINATION LEARNING DICTIONARY

Two sub-dictionaries with 2×12 kernels have been learned in the training part using combination learning and they are used for decomposing the test samples into sparse representation. In order to verify the high-efficency of the combination learning, time-comsuming and assessment results of a single dictionary with 1×24 kernels and the two sub-dictionaries will be compared. The initial single dictionary is designed to contain eight kernels in the three scales before learning and the learned single dictionary will be used for constructing the assessment index in the testing part, which can be seen

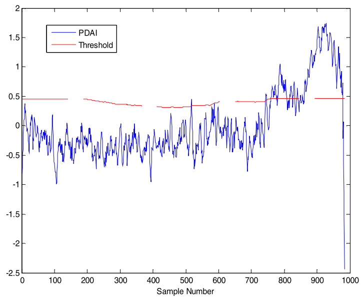

FIGURE 14. The fusion performance degradation index φ of single dictionary.

in Fig. 14. Compared with the assessment index based on combination learning dictionary and single dictionary, both of them have an increasing trend in the whole life. However, the index based on combination learning is more sensitive to fault degree because the description ability of different sub-dictionaries to the signal is different. The combination learning can select the best sub-dictionary for the new sample by comparing the assessment cost. On the other hand, the time consumptions of the two sub-dictionaries and the single dictionary are shown in Table 2. Obviously, the time-consuming of combination learning is about the half of single dictionary becauce the more initial kernels in the testing part will make the kernel seaching process more complex and the combination learning can compress the complexity exponentially. Thus, the learning learning can greatly improve the effectiveness of the proposed method.

TABLE 2. Time-consuming comparisons.

Dictionary type	Time Consumption (s)	
	One Sample	All Samples
Two Sub-dictionaries	3.9	3912
Single Dictionary	8.12	7946

V. CONCLUSIONS

A novel bearing performance degradation assessment method based on convolutional sparse combination learning is proposed in this paper. The convolutional sparse coding and ADMM are combined to extract the bearing vibration signal features without prior knowledge. The learned features have the ability to adapt to the complex operation conditions because the dictionaries are learned from the training samples. Moreover, the combination learning is integrated into the learning model to make the complexity of the model reduce greatly. A performance degradation assessment index is generated combining the error of the original and reconstructed signal with the kernel sparse norm, which can clearly describe the degradation tendency of the bearing in the whole life test. Compared with the traditional performance degradation assessment indexes, the indexes obtained by the proposed method is more sensitive to fault degree and more adaptive. Because of the excellent adaptive ability of the CSCL to the feature extraction of rotation machinery vibration signal, it can be used for remaining useful life prediction combined with time-series model and fault pattern recognition combined with classification model, which will be presented in the future.

REFERENCES

[1] Y. Wang, Y. Si, B. Huang, and Z. Lou, "Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017," *Can. J. Chem. Eng.*, vol. 96, no. 10, pp. 2073–2085, 2018.

[2] Y. Wang, Z. Hao, S. Wei, D. Zhou, and B. Huang, "Control performance assessment for ILC-controlled batch processes in a 2-D system framework," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 9, pp. 1493–1504, Sep. 2018.

[3] D. Djurdjanovic, J. Lee, and J. Ni, "Watchdog agent—An infotronics-based prognostics approach for product performance degradation assessment and prediction," *Adv. Eng. Inform.*, vol. 17, nos. 3–4, pp. 109–125, 2003.

[4] Z. Li, Y. Jiang, Q. Guo, C. Hu, and Z. Peng, "Multi-dimensional variational mode decomposition for bearing-crack detection in wind turbines with large driving-speed variations," *Renew. Energy*, vol. 116, pp. 55–73, Feb. 2018.

[5] Z. Li, D. Wu, C. Hu, and J. Terpenny, "An ensemble learning-based prognostic approach with degradation-dependent weights for remaining useful life prediction," *Rel. Eng. Syst. Saf.*, vol. 184, pp. 110–122, 2019.

[6] R. Yan and R. X. Gao, "Approximate entropy as a diagnostic tool for machine health monitoring," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 824–839, 2007.

[7] C. Sun, Z. Zhang, Z. He, Z. Shen, B. Chen, and W. Xiao, "Novel method for bearing performance degradation assessment—A kernel locality preserving projection-based approach," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 228, no. 3, pp. 548–560, 2014.

[8] C. Sun, Z. Zhang, Z. He, Z. Shen, and B. Chen, "Manifold learning-based subspace distance for machinery damage assessment," *Mech. Syst. Signal Process.*, vol. 70, pp. 637–649, Mar. 2016.

[9] Y. Pan, J. Chen, and L. Guo, "Robust bearing performance degradation assessment method based on improved wavelet packet–support vector data description," *Mech. Syst. Signal Process.*, vol. 23, no. 3, pp. 669–681, 2009.

[10] J. Yu, "Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models," *Mech. Syst. Signal Process.*, vol. 25, no. 7, pp. 2573–2588, 2011.

[11] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4137–4145, 2013.

[12] A. Widodo, B.-S. Yang, E. Y. Kim, A. C. C. Tan, and J. Mathew, "Fault diagnosis of low speed bearing based on acoustic emission signal and multi-class relevance vector machine," *Nondestruct. Test. Eval.*, vol. 24, no. 24, pp. 313–328, 2009.

[13] B. Eftekharnejad, M. R. Carrasco, B. Charnley, and D. Mba, "The application of spectral kurtosis on acoustic emission and vibrations from a defective bearing," *Mech. Syst. Signal Process.*, vol. 25, no. 1, pp. 266–284, 2011.

[14] F. Immovilli, A. Bellini, R. Rubini, and C. Tassoni, "Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison," *IEEE Trans. Ind. Appl.*, vol. 46, no. 4, pp. 1350–1359, Jul./Aug. 2010.

[15] A. Widodo *et al.*, "Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 7252–7261, 2009.

[16] W. Caesarendra, P. B. Kosasih, A. K. Tieu, A. S. Moodie, and B.-K. Choi, "Condition monitoring of naturally damaged slow speed slewing bearing based on ensemble empirical mode decomposition," *J. Mech. Sci. Technol.*, vol. 27, no. 8, pp. 2253–2262, 2013.

[17] K. Feng, Z. Jiang, W. He, and Q. Qin, "Rolling element bearing fault detection based on optimal antisymmetric real Laplace wavelet," *Measurement*, vol. 44, no. 9, pp. 1582–1591, 2011.
[18] W. Guo and P. W. Tse, "A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals," *J. Sound Vib.*, vol. 332, no. 2, pp. 423–441, 2013.
[19] P. Konar and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and support vector machines (SVMs)," *Appl. Soft Comput.*, vol. 11, no. 6, pp. 4203–4211, 2011.
[20] J. Seshadrinath, B. Singh, and B. K. Panigrahi, "Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets," *IEEE Trans. Power Electron.*, vol. 29, no. 2, pp. 936–945, Feb. 2014.
[21] W. Su, F. Wang, H. Zhu, Z. Zhang, and Z. Guo, "Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement," *Mech. Syst. Signal Process.*, vol. 24, no. 5, pp. 1458–1472, Jul. 2010.
[22] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[23] S. Mukherjee, R. Basu, and C. S. Seelamantula, "ℓ_1-K-SVD: A robust dictionary learning algorithm with simultaneous update," *Signal Process.*, vol. 123, pp. 42–52, Jun. 2016.
[24] R. Jiang, H. Qiao, and B. Zhang, "Efficient Fisher discrimination dictionary learning," *Signal Process.*, vol. 128, no. 1, pp. 28–39, 2016.
[25] A. M. Tillmann, "On the computational intractability of exact and approximate dictionary learning," *IEEE Signal Process. Lett.*, vol. 22, no. 1, pp. 45–49, Jan. 2015.
[26] D. L. Donoho, "For most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution," *Commun. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, 2010.
[27] G. He, K. Ding, and H. Lin, "Fault feature extraction of rolling element bearings using sparse representation," *J. Sound Vib.*, vol. 366, pp. 514–527, Mar. 2016.
[28] H. Liu, C. Liu, and Y. Huang, "Adaptive feature extraction using sparse coding for machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 25, no. 2, pp. 558–574, 2011.
[29] H. Tang, J. Chen, and G. Dong, "Sparse representation based latent components analysis for machinery weak fault detection," *Mech. Syst. Signal Process.*, vol. 46, no. 2, pp. 373–388, 2014.
[30] W. Fan, G. Cai, Z. K. Zhu, C. Shen, W. Huang, and L. Shang, "Sparse representation of transients in wavelet basis and its application in gearbox fault feature extraction," *Mech. Syst. Signal Process.*, vols. 56–57, pp. 230–245, May 2015.
[31] Y. Wang, J. Xiang, Q. Mo, and S. He, "Compressed sparse time–frequency feature representation via compressive sensing and its applications in fault diagnosis," *Measurement*, vol. 68, pp. 70–81, May 2015.
[32] H. Zhou, J. Chen, G. Dong, and R. Wang, "Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 65–79, May 2016.
[33] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, Jan. 2016.
[34] M. Šorel and F. Šroubek, "Fast convolutional sparse coding using matrix inversion lemma," *Digit. Signal Process.*, vol. 55, pp. 44–51, Aug. 2016.
[35] J. Ding, "Fault detection of a wheelset bearing in a high-speed train using the shock-response convolutional sparse-coding technique," *Measurement*, vol. 117, pp. 108–124, Mar. 2018.
[36] L. Yang *et al.*, "Image reconstruction via manifold constrained convolutional sparse coding for image sets," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1072–1081, Oct. 2017.
[37] J. Ding, F. Li, J. Lin, B. Miao, and L. Liu, "Fault detection of a wheelset bearing based on appropriately sparse impulse extraction," *Shock Vib.*, no. 71, pp. 1–17, 2017.
[38] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 391–398.
[39] R. Chalasani, J. C. Principe, and N. Ramakrishnan, "A fast proximal method for convolutional sparse coding," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2014, pp. 1–5.
[40] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2720–2727.
[41] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 217–224.
[42] Z. Li, J. Chen, Y. Zi, and J. Pan, "Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive," *Mech. Syst. Signal Process.*, vol. 85, pp. 512–529, 2017.
[43] Z. Feng, Y. Zhou, M. J. Zuo, F. Chu, and X. Chen, "Atomic decomposition and sparse representation for complex signal analysis in machinery fault diagnosis: A review with examples," *Measurement*, vol. 103, pp. 106–132, Jun. 2017.
[44] J. Lee, H. Qiu, G. Yu, and J. Lin. Rexnord Technical Services, 'Bearing Data Set'. IMS, University of Cincinnati, NASA Ames Prognostics Data Repository. NASA Ames, Moffett Field, CA, USA. [Online]. Available: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/
[45] 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 and 91, 2013.
[46] J. Zhou, H. Guo, L. Zhang, Q. Xu, and H. Li, "Bearing performance degradation assessment using lifting wavelet packet symbolic entropy and SVDD," *Shock Vib.*, pp. 1–10, 2016.

WENYI HUANG received the B.S. degree from East China Jiaotong University, Nanchang, China, in 2015. He is currently pursuing the Ph.D. degree with Hunan University, Changsha, China. His main research interests include pattern recognition and machinery fault diagnosis.

JUNSHENG CHENG received the Ph.D. degree in manufacturing engineering and automation from Hunan University, in 2005, where he is currently a Professor with the College of Mechanical and Vehicle Engineering. His main research interests include mechanical fault diagnosis, dynamics signal processing, vibration, and noise control.

YU YANG received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the College of Mechanical and Vehicle Engineering, Hunan University, Changsha, China, in 1994, 1997, and 2005, respectively. Her research interests include pattern recognition, digital signal processing, and machine fault diagnosis.