Sparse Representation Classification With Structured Dictionary Design Strategy for Rotating Machinery Fault Diagnosis

YUN KONG, TIANYANG WANG, ZHAOYE QIN, AND FULEI CHU
State Key Laboratory of Tribology, Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China

ABSTRACT Fault diagnosis technique is the core of Prognostics and Health Management (PHM) system, which plays a crucial role in the intelligent operation and maintenance of various rotating machineries. In this paper, we present a novel sparse representation classification framework with structured dictionary design strategy (SRC-SDD) for intelligent fault diagnosis of rotating machineries. The proposed SRC-SDD method consists of two stages, i.e., the structured dictionary design stage and the sparsity-based intelligent diagnosis stage. In the first stage, the novelty of SRC-SDD lies in the overlapping segmentation strategy for structured dictionary design, which leverages the structured prior knowledge of rotating machinery vibration signals, namely, the periodic self-similarity and shift-invariance properties. In the second stage, SRC-SDD achieves fault recognitions of testing samples using a sparsity-based diagnosis strategy based on the minimum sparse reconstruction error. The proposed structured dictionary design strategy can enhance the representation power of dictionaries and thus promote the recognition performance of the sparsity-based diagnosis strategy. Finally, the effectiveness of SRC-SDD has been validated on the gearbox fault dataset from IEEE PHM society. The diagnosis results show that SRC-SDD achieves the excellent recognition accuracy of 100% for predicting six different gearbox health states. Further, the comparative studies with three conventional SRC methods prove the superiority of SRC-SDD in terms of both the recognition performance and computation efficiency.

INDEX TERMS Fault diagnosis, periodic self-similarity, rotating machinery, sparse representation classification, structured dictionary design.

I. INTRODUCTION

Prognostics and health management (PHM) has been proven as one of the core technologies to promote the reliability and safety of various complex industrial systems in the era of smart manufacturing [1]–[4]. Fault diagnosis techniques play an important role in the PHM framework, which aim to early and accurately detect the component faults in the industrial equipment. The key components of rotating machinery such as the gears and bearings usually operate under the heavy and harsh working conditions, which are susceptible to the local damages and even lead to the catastrophic system failures. Therefore, it is of great significance to diagnosis the faults of rotating machineries early and accurately [5]. To early and accurately diagnose the rotating machinery faults, vibration-based signal analysis methods have been intensively studied and proven as one of the most effective techniques [6], which can be roughly categorized as the feature extraction methods and the pattern recognition methods.

Among the feature extraction methods, the classical time-domain statistics and spectral analysis techniques usually fail to detect the incipient weak faults due to the strong harmonic interferences and heavy background noises. To accurately extract and characterize the fault-related features from the complex multi-component signals of rotating machine, many advanced vibration-based signal processing algorithms have been developed, such as spectral...
In contrast to the fault feature extraction methods, pattern recognition methods can discover the underlying knowledge which implicitly represents various fault modes without the prior knowledge about the fault characteristic frequency. In recent years, the pattern recognition methods integrated with artificial intelligence technique have witnessed great success for rotating machinery fault diagnosis. Among the artificial intelligence techniques, deep learning (DL) has been proven one of the most popular options to enable pattern recognition methods for data-driven machinery fault diagnosis [21]–[25]. Wen et al. [22] incorporated the deep convolutional neural network (DCNN) with a signal-to-image conversion method for data-driven fault diagnosis. Shao et al. [23] improved the convolutional deep belief network (CDBN) with compressed sensing for efficient rolling bearing fault diagnosis. Sun et al. [24] developed the sparse deep stacking network (SDSN) method to overcome the overfitting risk of deep networks for motor fault diagnosis. Saufi et al. [25] proposed a stacked sparse autoencoder model for gearbox fault diagnosis in the case of limited sample data. Besides, the most recent works on deep transfer learning have attracted increasing attention to address the challenge of transferrable fault diagnosis of rotating machineries across different working conditions and even different machines. Although these DL techniques enable to learn the abstract features for data-driven machine fault diagnosis, the diagnosis performances of DL techniques still heavily rely on the well-designed complex deep network architectures and various finely-tuned hyperparameters [26].

Alternatively, the emerging sparse representation-based classification (SRC) methods have also been widely applied for computer vision and pattern recognition [27], [28], which enable pattern recognition with sparse representation theory. Compared with DL techniques, SRC methods do not require designing complex deep network architectures and the labor-intensive parameter tuning works, which can be regarded as a promising tool for rotating machine fault diagnosis as well. The core idea of sparse representation theory is to represent or approximate the signal of interest \( y \in \mathbb{R}^n \) over the so-called redundant dictionary \( D \in \mathbb{R}^{n \times K} \) using a linear combination of few atoms (columns of the dictionary \( D \)), i.e., \( y = Dx \) or \( y \approx Dx \), where \( x \in \mathbb{R}^K \) is the sparse code of \( y \) containing just a few non-zero coefficients. The SRC methods for pattern recognition can be roughly categorized as the reconstruction error-based SRC and the classifier training-based SRC methods. Within these SRC methods, the dictionary can be designed by either using the training samples directly [29] or using dictionary learning algorithms such as K-SVD (K-means singular value decomposition) [30]. On the one hand, the core idea of the reconstruction error-based SRC is that testing samples tend to achieve the minimum reconstruction error with respect to the category-specific dictionary designed using the training samples with the same class label [29]. Inspired by this idea, Tang et al. [31] achieved the sparse classification of rolling bearing faults using the reconstruction error-based SRC with compressive sensing strategy. The dictionary learning-based sparse classification approaches were developed for planet bearing fault diagnosis [32] and epilepsy detection [33]. Han et al. [34] incorporated K-SVD for dictionary learning with the reconstruction error-based SRC method for wind turbine fault diagnosis. Later, Wang et al. [35] further proposed the weighted sparse representation classification method using an improved K-SVD algorithm for bearing fault diagnosis. On the other hand, the core idea of the classifier training-based SRC method is incorporating the reconstruction power of signal sparse representation and the discrimination power of classifiers, which is achieved by using the sparse codes as input features to train the classifier for pattern recognition. To this end, Zhang and Li [36] proposed the discriminative K-SVD algorithm to learn the dictionary and classifier jointly for face recognition. Jiang et al. [37] further developed a Fisher discriminative K-SVD method by imposing the Fisher discrimination criterion on sparse codes for face recognition. Aiming to obtain both the reconstructive and discriminative power from dictionary learning, Zheng and Tao [38] proposed a label consistent K-SVD method to learn the discriminative dictionary and an optimal linear classifier simultaneously for pattern recognition. Ren et al. [39] developed a new multi-view SRC method based on joint supervised dictionary and classifier training for synthetic aperture radar (SAR) image classification. Recently, Kong et al. [40] developed the discriminative dictionary learning based sparse classification method for planet bearing fault diagnosis. However, when applying these above SRC methods for rotating machinery fault diagnosis, the structured prior knowledge of rotating machinery vibration signals are not considered to enhance the discriminative power of SRC methods and thus may lead to the inferior fault recognition performances.
and fault recognition stage. The main contributions of this paper are presented as follows.

1) In the structured dictionary design stage, the periodic self-similarity and shift-invariance properties of rotating machinery vibration signals are fully considered using an overlapping segmentation strategy, which can enable to design the category-specific dictionaries for each health state of rotating machines and enhance the discriminative power of SRC for accurate fault recognitions.

2) In the fault recognition stage, a sparsity-based intelligent diagnosis strategy is established using the discrimination criterion based on minimum sparse reconstruction error of testing samples with respect to all structured category-specific dictionaries.

3) Compared with some state-of-the-art SRC methods, the experiment validation results on the famous gearbox fault dataset from IEEE PHM Society demonstrate that SRC-SDD achieves the highest recognition performance and lowest computation costs for machine fault diagnosis.

The rest of this paper is organized as follows. Section II presents the related works on the sparse representation-based classification (SRC) methods and the limitations for rotating machinery fault diagnosis. In Section III, we introduce SRC-SDD for rotating machinery fault diagnosis, including the structured dictionary design strategy, and the sparsity-based intelligent diagnosis strategy. The experiment validations of SRC-SDD for rotating machine fault diagnosis are presented in Section IV. Finally, Section V draws the conclusions.

II. RELATED WORKS

A. SPARSE REPRESENTATION AND DICTIONARY LEARNING

Different from the classical signal representation paradigms using complete orthogonal bases (such as Fourier transform), sparse representation has brought new insights into the signal representation paradigm using redundant dictionaries. The basic idea of sparse representation is to reveal the sparsest way to represent or approximate the input signal by a linear combination of few dictionary atoms (columns of dictionary), as illustrated in Fig. 1(a). Given that \( y \in \mathbb{R}^n \) denotes the input signal and \( D = [d_1, d_2, \ldots, d_K] \in \mathbb{R}^{n \times K} \) stands for the dictionary, the sparse code \( x \in \mathbb{R}^K \) (i.e., sparse representation coefficients) of \( y \) with respect to \( D \) can be calculated by solving the sparse coding problem as follows,

\[
x^* = \arg \min_{x} \left\| y - Dx \right\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq T,
\]

where the \( L_0\)-norm \( \|x\|_0 \) defines the sparsity constraint and \( T \) is the sparsity threshold to enforce the number of nonzero entries in the vector \( x \). Typical sparse coding algorithms for solving (1) are the sparse regularization and greedy pursuit methods. On the one hand, the sparse regularization methods solve the following least squares problem regularized with a sparsity-inducing penalty,

\[
x^* = \arg \min_{x} \frac{1}{2} \left\| y - Dx \right\|_2^2 + \lambda \phi(x), \quad \lambda > 0,
\]

where \( \lambda \) is the regularization parameter and \( \phi(x) \) is a sparsity-inducing penalty such as the convex \( L_1 \)-norm and nonconvex penalties including the arctangent penalty [17], generalized minimax-concave penalty [18]. The optimization algorithms to solve the sparse regularization problem in (2) include the proximal splitting method [41], and the alternating direction method of multipliers [42]. On the other hand, the greedy pursuit methods solve the sparse coding problem in (1) in a greedy way to get the approximate solution satisfying the \( L_0 \)-norm constraint. Among the greedy pursuit algorithms, the orthogonal matching pursuit (OMP) is a strong competitor due to its computation efficiency for sparse coding [43]. The procedure of OMP for sparse coding is listed in Algorithm 1.

Dictionary plays an important role in the effective sparse representations, which is hard to predefine due to the limited prior knowledge about input signals. In this case, dictionary learning method is helpful, which learns to design dictionary for sparse representation using the optimization models and training signals. Assume that \( Y = [y_1, y_2, \ldots, y_N] \in \mathbb{R}^{N \times n} \) denotes \( N \)-dimensional training samples, dictionary learning solves the following optimization problem to learn the data-driven dictionary \( D^* \).

\[
[D^*, X^*] = \arg \min_{D,X} \left\| Y - DX \right\|_F^2 = \sum_{i=1}^{N} \left\| y_i - Dx_i \right\|_2^2
\]

s.t. \( \forall i, \|x_i\|_0 \leq T \),

where \( D = [d_1, d_2, \ldots, d_K] \in \mathbb{R}^{n \times K} \) is the dictionary containing \( K \) atoms (\( K \) is the dictionary size) and \( X = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{K \times N} \) is the sparse codes of \( Y \) with respect to dictionary \( D \). The basic principle of dictionary learning is shown in Fig. 1(b). The typical dictionary learning algorithm
Algorithm 1 OMP for Sparse Coding

**Input:** Input signal $y \in \mathbb{R}^n$, dictionary $D=[d_1, d_2, \ldots, d_k] \in \mathbb{R}^{n \times K}$, and sparsity threshold $T$.

**Initialize:** Residual $r_0 = y$, index set $A_0 = \emptyset$, the activated atom set $D_0 = \emptyset$, and iteration number $t = 1$.

**Procedure:**
1. for $t = 1, 2, \ldots, T$, do
2. Find the index $y_t$ that solves the optimization problem,
   $$y_t = \arg \max_{k=1, \ldots, K} |\langle r_{t-1}, d_k \rangle|.$$
3. Update the index set $A_t = A_{t-1} \cup \{y_t\}$ and the activated atom set $D_t = [D_{t-1}, d_{y_t}]$.
4. Find the approximate solution of a least squares problem,
   $$\hat{x}_t = \arg \min_{\hat{x}} \| y - D_t \hat{x} \|_2.$$
5. Update the residual $r_t = y - D_t \hat{x}_t$ and the iteration number $t = t + 1$.
6. end for

**Output:** Sparse code $\hat{x}_T$.

is K-SVD [30], which alternatively implements the sparse coding and dictionary update procedures to minimize the cost function in (3).

### B. SPARSE REPRESENTATION-BASED CLASSIFICATION

In recent years, the sparse representation-based classification (SRC) methods have attracted lots of attention for computer vision and pattern recognition tasks [27]. Based on different discrimination criteria for classifications, SRC methods are categorized into the reconstruction error-based SRC and the classifier training-based SRC.

1) RECONSTRUCTION ERROR-BASED SRC

The basic idea of the reconstruction error-based SRC is that the testing signals tend to be sparsely approximated and achieve the minimum reconstruction error with respect to the dictionary designed using training signals from the correct category [29], [32]. Given $L$ sets of training signals from $L$ different categories, the reconstruction error-based SRC methods achieve pattern recognition via the following steps.

**Step 1:** Design $L$ different category-specific dictionaries $D_l$ ($l = 1, 2, \ldots, L$) using the training signals directly [29] or the dictionary learning algorithm [32] for each category.

**Step 2:** Calculate the sparse codes $x^*_l$ ($l = 1, 2, \ldots, L$) of the testing signal $y$ with respect to $L$ different category-specific dictionaries $D_l$ by sparse coding in (1) and the corresponding reconstruction errors $RE(y, D_l, x^*_l)$ as follows,

$$RE(y, D_l, x^*_l) = \| y - D_l x^*_l \|_2^2, \quad l = 1, 2, \ldots, L. \quad (4)$$

**Step 3:** Predict the class label of $y$ using the discrimination criterion based on the minimum reconstruction error,

$$\text{label}(y) = \arg \min_l \ RE(y, D_l, x^*_l), \quad l = 1, 2, \ldots, L. \quad (5)$$

2) CLASSIFIER TRAINING-BASED SRC

The other strategy to gain discrimination power for SRC methods is to incorporate dictionary learning with classifier training. The basic idea of the classifier training-based SRC methods is to optimize the classifier model for pattern recognition by exploiting sparse codes of the training signals as inputs of the classifiers [36]–[40]. Given the training signals $Y \in \mathbb{R}^{n \times N}$ and the associated label matrix $H=[h_1, h_2, \ldots, h_N] \in \mathbb{R}^{L \times N}$, the classifier training-based SRC methods achieve pattern recognition as follows.

**Step 1:** Learn the discriminative dictionary and classifier model jointly by solving a unified optimization problem,

$$\{D^*, W^*, X^*\} = \arg \min_{D, W, X} \| Y - DX \|_F^2$$

$$+ \lambda_1 \sum_{i=1}^N C \{ h_i, f(x_i, W) \} + \lambda_2 \| W \|_F^2,$$

s.t. $\forall i, \| x_i \|_0 \leq T, \quad (6)$

where $D \in \mathbb{R}^{n \times K}$ and $W \in \mathbb{R}^{K \times K}$ represents the discriminative dictionary and the classifier model parameters, respectively. $\lambda_1$ and $\lambda_2$ are the regularization parameters. Besides, $C$ is the classification loss and $f(x_i, W)$ is the predicted label using the sparse code $x_i$ as input features of the classifier $W$.

**Step 2:** Calculate the sparse code $x^*$ of the testing signal $y$ with respect to the optimized discriminative dictionary $D^*$,

$$x^* = \arg \min_x \| y - D^* x \|_2^2 \quad \text{s.t.} \quad \| x \|_0 \leq T. \quad (7)$$

**Step 3:** Assign the class label of the testing signal $y$ as the index corresponding to the largest element in $\text{label}(y)$, which is estimated using the optimized $W^*$ and the sparse code $x^*$,

$$\text{label}(y) = W^* x^* \quad (8)$$

### C. LIMITATIONS OF THE EXISTING SRC METHODS

When applying these existing SRC methods for rotating machinery fault diagnosis, there are still some limitations to be addressed. First, the recognition performances of SRC methods rely on the representation capability of dictionaries. However, the dictionary design strategy in the existing SRC methods neglects the structured prior knowledge of rotating machinery vibration signals, which may limit the recognition performance of SRC for rotating machinery fault diagnosis.

Second, the local features between the training samples and testing samples are not fully considered in the existing SRC methods [36], which plays a key role in the SRC methods for pattern recognition. Therefore, it is promising to develop a wise dictionary design strategy leveraging the structured prior knowledge of rotating machinery vibration signals and the helpful similarity information to promote the recognition performances of SRC for rotating machinery fault diagnosis.

### III. PROPOSED SRC-SDD METHOD

The proposed sparse representation classification method with structured dictionary design strategy (SRC-SDD) is presented for rotating machinery fault diagnosis in this
section, which involves two stages, i.e., the structured dictionary design stage and the sparsity-based intelligent diagnosis stage.

A. STRUCTURED DICTIONARY DESIGN

To overcome the limitations of the existing SRC methods for rotating machinery fault diagnosis, we propose the structured dictionary design strategy to leverage the key structured prior knowledge of rotating machinery vibration signals, namely, the periodic self-similarity and shift-invariance properties, for enhancing the representation power of dictionary and the fault recognition performances of SRC methods.

1) PERIODIC SELF-SIMILARITY OF ROTATING MACHINERY VIBRATION SIGNALS

When rotating machinery operates under constant working conditions, the repetitive working modes of the rotating machinery will produce the vibration response signals with strong periodicity, thus determining the periodic self-similarity of rotating machine vibration signals. Taking the faulty vibration signals of rotating machinery such as the gear transmission systems as an example, different rotating parts can generate multiple vibration components with strong periodic self-similarity, as illustrated in Fig. 2(a). The faulty vibration signals of rotating machinery generally consist of four constituent components, namely, the harmonic features, amplitude modulation (AM) and frequency modulation (FM) components, repetitive transients and random noises. Among the first three constituent components, they both share strong periodic self-similarity property, such as harmonic features induced by rotating shafts and healthy meshing gears, AM-FM components induced by the distributed gear faults, and repetitive transients induced by the localized bearing defects.

Therefore, identifying the periodic self-similarity property of rotating machinery vibration signals, especially the fault-related feature components, is crucial for rotating machinery fault diagnosis under constant working conditions [44]. As for the SRC methods, leveraging the periodic self-similarity property of rotating machine vibration signals for dictionary design under different health states will obviously enhance the representation capability of dictionary and promote the fault recognition performances.

2) SHIFT-INVARIANCE PROPERTY WHEN PREDICTING HEALTH STATES

During the process of rotating machinery condition monitoring using sensors (such as accelerometers), the data acquisition system will acquire large amount of sensory data with the sampling time increasing. In the real applications of rotating machinery fault diagnosis using pattern recognition methods, the long-term monitored sensory data will be first divided into many short data segments and then these short data segments will be used as inputs of pattern recognition methods for health state prediction. As illustrated in Fig. 2(b), if the rotating machinery operates under a certain health state and the degradation of health state can be negligible during a short time, then a large amount of short data segments will correspond to different time history but all indicate the same health state. In this case, the health state prediction results of these short data segments by the same pattern recognition method should be ideally invariant to the time development.
the time-shifts of these short data segments. In other words, if we applied the pattern recognition methods for rotating machinery fault diagnosis, the predicted health state labels of these short data segments during a short time should be ideally consistent with the only real health state, i.e., satisfying the shift-invariance property.

More specifically, as shown in Fig. 2(b), we consider the rotating machinery operating under the certain health state \( l \), the monitored vibration data \( y_{\text{state} \ l} \) by sensors can be divided into a large amount of short data segments, which correspond to different time history but all indicate the same health state \( l \). In this case, we assume that there are \( N \) short data segments \( \{y_1, y_2, \ldots, y_N\} \) indicating the same health state \( l \), the predicted health state label results of two different short data segments \( y_x \) and \( y_y \) by the same classifier \( W \) should be ideally consistent with the health state \( l \) and invariant to the time-shifts of \( y_x \) and \( y_y \) in the whole vibration data \( y_{\text{state} \ l} \). In other words, the predicted health state labels of the short data segments \( y_x \) and \( y_y \) satisfy the following shift-invariance property,

\[
\forall s, t \in \{1, 2, \ldots, N\}, \quad \text{label}(y_t, W) = \text{label}(y_s, W),
\]

where \( \text{label}(y_x, W) \) and \( \text{label}(y_y, W) \) are the predicted health state of the short data segment \( y_x \) and \( y_y \) using the classifier \( W \), respectively. It should be mentioned that the short data segments \( y_x \) and \( y_y \) only differ in the different starting points of the vibration data \( y_{\text{state} \ l} \) under the health state \( l \).

3) STRUCTURED DICTIONARY DESIGN USING OVERLAPPING SEGMENTATION STRATEGY

Proper dictionary design strategy is important to represent the structured prior knowledge of rotating machinery vibration signals, namely, the periodic self-similarity and shift-invariance property. Exploiting the periodic self-similarity and shift-invariance properties for structured dictionary design can enhance the representative power of dictionaries for effective sparse representations and promote the discriminative power of the sparsity-based fault diagnosis methods. To this end, we propose an overlapping segmentation strategy for the structured dictionary design as follows, which uses the raw rotating machine vibration data.

The overlapping segmentation strategy for the structured category-specific dictionary design is shown in Fig. 3. Firstly, we define a segmentation operator \( R_k \) which is parameterized with two segmentation parameters, i.e., the window size \( W \) and overlapping rate \( \delta \). Different segmentation operations for 1-D mechanical vibration signals can be achieved using the segmentation operator \( R_k \) with different parameters \( W \) and \( \delta \).

As illustrated in Fig. 3(a), assume that \( y \in \mathbb{R}^{1 \times m} \) denotes the 1-D mechanical vibration signal, the segmentation operator \( R_k: \mathbb{R}^{1 \times m} \rightarrow \mathbb{R}^{W \times 1} \) is defined to extract the transpose of the \( k \)-th local data segment \( y_k \) of the 1-D mechanical vibration signal \( y \) and form as the \( k \)-th atom of dictionary \( D \) as follows,

\[
R_k(y) = y_k^T = d_k = D(:, k).
\]

where \( d_k \) is the \( k \)-th atom in the dictionary \( D \). Further, the overlapping segmentation operator \( R \) is defined to transform the 1-D vibration signal \( y \) into a 2-D dictionary matrix. In specific, the overlapping segmentation operator \( R \) consists of \( K \) cascaded segmentation operator, i.e., \( R = [R_1, R_2, \ldots, R_{K-1}, R_K] \) \((K \text{ is the dictionary size})\). As a result, as illustrated in Fig. 3(b), the overlapping segmentation operator \( R: \mathbb{R}^{1 \times m} \rightarrow \mathbb{R}^{W \times K} \) designs the structured 2-D dictionary \( D \in \mathbb{R}^{W \times K} \) using the raw 1-D mechanical vibration signal \( y \) directly as follows,

\[
D = [R_1(y), R_2(y), \ldots, R_{K-1}(y), R_K(y)].
\]

Moreover, if we consider \( L \) different health states of rotating machinery and assume that \( y_l \in \mathbb{R}^{1 \times m} \) stands for the 1-D raw vibration data under the health state \( l \), then we can apply the overlapping segmentation operator \( R \) to design \( L \) different structured category-specific dictionaries \( D_l \) for each health state \( l \) \((l = 1, 2, \ldots, L)\) using the vibration data \( y_l \) as follows,

\[
\forall l = 1, 2, \ldots, L, \quad D_l = R(y_l) = [R_1(y_l), R_2(y_l), \ldots, R_{K-1}(y_l), R_K(y_l)].
\]

In summary, the structured dictionary design for \( L \) different health states using the overlapping segmentation strategy is illustrated in Stage I of Fig. 4. It is worth noting that the structured dictionary design strategy could leverage the local features between testing samples and testing samples, since the overlapping segmentation operator maximally maintains the self-similarity of local data segments for each health state.

It should be mentioned that the overlapping segmentation strategy for the structured dictionary design also applies to the raw testing vibration signals, which could generate plenty of testing samples with unknown health states for health state prediction in the second stage of the SRC-SDD method.
Algorithm 2 SRC-SDD for Rotating Machine Fault Diagnosis

**Input:** Training signal set \( \{ y_l \} (l = 1, 2, \ldots, L) \), window size \( W \), overlapping rate \( \delta \), sparsity threshold \( T \) and testing sample \( y_i \).

**Procedure:**
1: for \( l = 1, 2, \ldots, L \), do
2: Design the category-specific dictionary \( D_l \) via (10)-(12).
3: end for
4: for \( l = 1, 2, \ldots, L \), do
5: Compute sparse codes \( \hat{x}_i \) of testing sample \( y_i \) with respect to \( L \) category-specific dictionary \( D_l \) via (13) using OMP.
6: Compute the reconstruction errors \( \text{RE}(y_i, D_l, \hat{x}_i^{(l)}) \) via (14).
7: end for
8: Predict the health state \( \text{label}(y_i) \) via (15).

**Output:** The health state \( \text{label}(y_i) \) of testing sample \( y_i \).

2) **RECONSTRUCTION ERRORS CALCULATION**

As for the testing sample \( y_i \), we then calculate the sparse reconstruction errors \( \text{RE}(y_i, D_l, \hat{x}_i^{(l)}) \) of \( y_i \) using the sparse approximations of \( y_i \) with respect to the \( L \) different category-specific dictionaries \( D_l \) for each health state \( l \) \( (l = 1, 2, \ldots, L) \) as follows,

\[
\text{For } l = 1, 2, \ldots, L, \quad \text{RE}(y_i, D_l, \hat{x}_i^{(l)}) = \|y_i - D_l \hat{x}_i^{(l)}\|_2^2.
\]  

(14)

3) **HEALTH STATE LABEL PREDICTION**

Thirdly, as for the testing sample \( y_i \), we predict the health state \( \text{label}(y_i) \) of \( y_i \) according to the following discrimination criterion based on minimum sparse reconstruction error,

\[
\text{label}(y_i) = \arg \min_{l=1,2,\ldots,L} \text{RE}(y_i, D_l, \hat{x}_i^{(l)}).
\]  

(15)

The above sparsity-based intelligent diagnosis strategy for fault recognition assumes that samples with the same health state label tend to share similar sparse representations. Thus, testing samples are highly possible to achieve the minimum reconstruction errors using the sparse approximations with respect to the structured category-specific dictionary, which is designed using training signals from the correct category.

In summary, the algorithmic procedures of SRC-SDD for rotating machine fault diagnosis are detailed in Algorithm 2.

**IV. EXPERIMENT VALIDATIONS**

In this section, the effectiveness and superiority of the SRC-SDD method for rotating machinery fault diagnosis are evaluated on the parallel-shaft gearbox fault dataset, which was released in the 2009 data challenge by IEEE Prognostics and Health Management (PHM) Society. Moreover, the fault diagnosis results of the state-of-the-art sparse representation-based classification (SRC) methods including the dictionary learning based SRC (DL-SRC) [32] and the label consistent...
KSVD (LC-KSVD1 and LC-KSVD2) [37] are compared in terms of both the recognition performance and computation efficiency. Finally, the effects of algorithm parameters on the recognition performance of SRC-SDD are analyzed in depth.

A. CASE STUDY: GEARBOX FAULT DIAGNOSIS

1) EXPERIMENT DESCRIPTION

The gearbox fault dataset released by IEEE PHM Society [45] is collected on a generic industrial gearbox, which consists of three shafts, four spur or helical gears and six bearings, as shown in Fig. 5. As for the signal acquisitions, two accelerometers installed on the input and output shaft retaining plates and a tachometer are used to acquire the vibration data and rotational speed of the input shaft, respectively. The vibration signals of spur gears acquired by the accelerometer on the input side are analyzed in this case. The specific fault modes on this gearbox dataset contain three gear fault modes, three bearing fault modes, the shaft imbalance, the normal gear, and the normal bearing, as illustrated in Fig. 6. In this case study, six different gearbox health states are considered and their fault modes are detailed in Table 1. During the vibration measurement experiments, the rotation speed of the input shaft is 30 Hz and the gearbox operates with a high load. Besides, the sampling rate is 200/3 kHz and the sampling time of each vibration acquisition is 4 seconds. Two different vibration measurements are acquired under each gearbox health state, one of which serves as the training dataset for algorithm training and the other serves as the testing dataset for fault recognition. As a result, both the training and testing vibration signals of six different gearbox health states are illustrated in Fig. 7.

2) FAULT RECOGNITION RESULTS BASED ON SRC-SDD

In this case study, algorithm parameters of SRC-SDD for fault recognition of gearbox health states are set as follows. First, as for the structured dictionary design using the overlapping segmentation strategy, the window size $W$ and overlapping rate $\delta$ are selected as 6250 and 0.97, respectively. As a result, the number of training data segments for each gearbox health state is 1394 and the dictionary size $K$ of the class-specific dictionary $D_{cl}$ for each gearbox health state is 1394 as well. Second, as for the sparsity-based diagnosis strategy in the fault recognition stage, the sparsity threshold $T$ for sparse coding is selected as 10. Third, as for the testing samples, the overlapping segmentation strategy with the same parameters $W$ and $\delta$ for structured dictionary design is used to generate the testing samples. Therefore, each gearbox health state has 1394 testing samples and the total testing samples for health state prediction is 8364. Finally, the label prediction results of SRC-SDD on the gearbox fault dataset are shown in Fig. 8. It can be observed from Fig. 8 that there is no misclassified testing sample among the total 8364 samples and SRC-SDD always achieves the excellent recognition accuracy of 100% for fault recognition of all the six different gearbox health states. These above health state recognition results prove the effectiveness of SRC-SDD for gearbox fault diagnosis.

3) COMPARATIVE STUDIES

To demonstrate the superiority of SRC-SDD for rotating machine fault diagnosis, the fault recognition results of SRC-SDD are compared with the state-of-the-art SRC methods. As for the comparative methods, we select DL-SRC, LC-KSVD1, and LC-KSVD2. DL-SRC [32] exploits the KSVD-based dictionary learning algorithm for dictionary design and achieves the fault recognition based on minimum reconstruction error for fault recognition of planet bearings. Moreover, LC-KSVD1 and LC-KSVD2 [37] have been proven one of the most successful SRC methods, which learns the discriminative dictionary and an optimal linear classifier model jointly for pattern recognition. However,
FIGURE 7. Training and testing vibration signals of six different health states on the gearbox fault dataset. (a)-(b) Health state 1, (c)-(d) Health state 2, (e)-(f) Health state 3, (g)-(h) Health state 4, (i)-(j) Health state 5, (k)-(l) Health state 6.

FIGURE 8. Label prediction results of SRC-SDD for gearbox fault diagnosis.

FIGURE 9. Confusion matrices of different methods for gearbox fault diagnosis. (a) SRC-SDD, (b) DL-SRC, (c) LC-KSVD1, and (d) LC-KSVD2.

the above three SRC methods all neglect the structured prior knowledge of rotating machine vibration signals, which may decrease the fault recognition performances of these methods.

As for the comparative studies, both the fault recognition performance and the computation efficiency results of SRC-SDD and three comparative methods (DL-SRC, LC-KSVD1, and LC-KSVD2) are presented as follows.

Fig. 9 illustrates the confusion matrices of SRC-SDD and three comparative SRC methods for gearbox fault diagnosis. We can observe that SRC-SDD obtains the best recognition performance, while three comparative methods (DL-SRC, LC-KSVD1, and LC-KSVD2) all misclassify several testing samples on this gearbox fault dataset. In specific, DL-SRC achieves the recognition accuracies of 0.84, 0.91, 0.46, 0.69, 0.70 and 0.74 for identifying the six different gearbox health states, respectively. Moreover, the recognition accuracies of LC-KSVD1 for identifying six gearbox health states are 0.87, 0.92, 1, 0.93, 0.76 and 1, respectively. As for LC-KSVD2, it obtains the recognition accuracies of 0.85, 0.92, 1, 0.97, 0.79 and 1 for fault recognition of six health states, respectively.

For the purpose of quantitively comparing the recognition performance of different methods for gearbox fault diagnosis, both the overall recognition accuracies and F1-score of SRC-SDD and the comparative SRC methods are calculated. The F1-score is a useful indicator for quantitively evaluating the performances of the pattern recognition approaches, which comprehensively considers the precision rate and the recall rate [46]. F1-score is the harmonic mean of the precision rate $P$ and the recall rate $R$, which is defined as follows,

$$P = \frac{TP}{TP + FP},$$  \hspace{1cm} (16)  

$$R = \frac{TP}{TP + FN},$$  \hspace{1cm} (17)  

$$\text{F1-score} = \frac{2 \times P \times R}{P + R},$$  \hspace{1cm} (18)  

where the precision rate $P$ is the ratio of the number of true positives ($TP$) to the sum of $TP$ and the number of false positives ($FP$), and the recall rate $R$ is the ratio of $TP$ to the sum of $TP$ and the number of false negatives ($FN$). The F1-score reaches the worst value at 0 and the best value at 1.
Fig. 10 displays quantitative recognition results of different SRC methods for gearbox fault diagnosis. The comparison results of recognition accuracy are shown in Fig. 10(a), which indicates that SRC-SDD always obtains the best recognition accuracies for predicting six gearbox health states. Moreover, the overall average recognition accuracy for SRC-SDD is 100%, which is superior to the overall average recognition accuracies of DL-SRC (72.22%), LC-KSVD1 (91.34%), and LC-KSVD2 (92.13%) for gearbox fault diagnosis. Further, the comparison results of F1-score are shown in Fig. 10(b), which show that SRC-SDD also always achieves the best F1-score for six different gearbox health states and the highest overall average F1-score (100%) on the gearbox fault dataset. In contrast, DL-SRC, LC-KSVD1 and LC-KSVD2 gain the overall average F1-score of 72.05%, 91.33%, and 92.06%, respectively. The above quantitative recognition results prove that SRC-SDD outperforms three comparative SRC methods for reliable gearbox fault diagnosis.

Finally, the computation efficiency of different methods are quantitatively compared. The computation time results of SRC-SDD and the comparative SRC methods on the gearbox fault dataset are shown in Fig. 11. It should be noted that all these four methods are executed under Windows 7 operating system and the MATLAB 2016b environment on a computer equipped with the Intel Xeon CPU E5 at 2.2 GHz and RAM of 64 GB. We can observe from Fig. 11 that SRC-SDD costs the least training time (3.25 s), which is far superior to the training time of DL-SRC (524.18 s), LC-KSVD1 (1390.57 s) and LC-KSVD2 (1287.81 s). Besides, as for the testing time for predicting health states of the total 8364 testing samples, SRC-SDD costs 20.68 seconds, which is competitive to the testing time of DL-SRC (24.74 s) but inferior to the testing time of LC-KSVD1 (0.75 s) and LC-KSVD2 (0.51 s). In total, the total computation time considering both the training and testing of SRC-SDD is 23.93 s, which is the most efficient among the total computation time of four different methods. In contrast, the total computation time of the DL-SRC, LC-KSVD1 and LC-KSVD2 are 548.92, 1391.32 and 1288.32 seconds, respectively. The reason for the superiority of SRC-SDD is that SRC-SDD exploits the structured prior knowledge of rotating machinery vibration signals directly for dictionary designs, while DL-SRC, LC-KSVD1 and LC-KSVD2 both apply the time-consuming dictionary learning algorithms to design dictionaries. These above computation time results prove the superiority of SRC-SDD in terms of computation efficiency for gearbox fault diagnosis.

B. PARAMETER ANALYSIS

The algorithm parameters play an important role in the fault recognition performance of SRC-SDD for rotating machinery fault diagnosis. In this section, the effects of three algorithm parameters on the recognition accuracies of SRC-SDD for gearbox fault diagnosis are thoroughly investigated using the cross-validation strategy.

1) OVERLAPPING SEGMENTATION PARAMETERS $W$ AND $\delta$

The window size $W$ and overlapping rate $\delta$ are two crucial parameters in the structured dictionary design stage using the overlapping segmentation strategy. The parameters $W$ and $\delta$ influence the dimensions of the category-specific dictionary for each gearbox health state and affect the representative power of dictionaries in the sparsity-based diagnosis strategy. The structured dictionary design strategy with a greater value of the overlapping rate $\delta$ could take better advantage of the structured prior knowledge of rotating machinery vibration signals and thus could promote the recognition performance of SRC-SDD. To validate this implication, the effects of the window size $W$ and overlapping rate $\delta$ on the recognition accuracy of SRC-SDD are investigated using the grid search method, and the obtained results are shown in Fig. 12. Here, the window size $W$ is varied within the range of [1000,10000] with a search step of 250, and the overlapping rate $\delta$ is varied within the search range of [0.2:0.2:0.8, 0.9, 0.93:0.02:0.97].
According to the recognition accuracy results in Fig. 12, we could reach the following conclusions. First, SRC-SDD with the greater overlapping rate $\delta$ leads to a higher recognition accuracy than SRC-SDD with a small overlapping rate $\delta$ for gearbox fault diagnosis. Second, in the cases using the great overlapping rate $\delta$ (such as 0.97), the recognition accuracy increases slowly with the increasing of the window size $W$. While in the cases using the small overlapping rate $\delta$ (such as 0.2, 0.4 and 0.6), the recognition accuracy decreases with several fluctuations when increasing the window size $W$. Based on these observations, in the general tasks of rotating machinery fault diagnosis, the relatively great overlapping rate is recommended to choose to exploit the structured prior knowledge of rotating machinery vibration signals as much as possible, and the window size should be properly selected using the cross validation strategy.

2) SPARSITY THRESHOLD $T$

The sparsity threshold $T$ is the crucial parameter of OMP algorithm for sparse coding in the sparsity-based diagnosis strategy of SRC-SDD. The effect of sparsity threshold $T$ on the recognition accuracy of SRC-SDD for gearbox fault diagnosis is shown in Fig. 13, where the sparsity threshold $T$ is varied within the range of [5:1:60]. Base on the recognition accuracy results in Fig. 13, we can clearly observe that the recognition accuracies of SRC-SDD are insensitive to the variations of the sparsity threshold $T$. The subplot in Fig. 13 further indicates that the variations of the recognition accuracy of SRC-SDD with different choices of the sparsity threshold $T$ are negligible and the optimal sparsity threshold $T$ to gain the best recognition performance of SRC-SDD is 10 for gearbox fault diagnosis.

V. CONCLUSION

In this paper, a new sparse representation-based classification method with structured dictionary design strategy (SRC-SDD) is proposed for rotating machinery fault diagnosis. The SRC-SDD method consists of the dictionary design stage and the sparsity-based intelligent diagnosis stage. The structured prior knowledge of rotating machinery vibration signals (periodic self-similarity and shift-invariance) are exploited to achieve structured dictionary designs by the overlapping segmentation strategy. The proposed structured dictionary design strategy could also leverage the similarity features between the labeled training samples and the unlabeled testing samples for robust health state predictions. Besides, the sparsity-based intelligent diagnosis strategy is established according to a discrimination criterion based on the minimum sparse reconstruction error, which could achieve robust fault recognition by the merits of the strong representative power of the designed structured sub-dictionaries for each health state. Finally, the effectiveness and superiority of SRC-SDD have been validated on the gearbox fault dataset from IEEE PHM Society. The diagnosis results show that SRC-SDD obtains the excellent overall recognition accuracy of 100% for intelligent identification of six different gearbox health states. Besides, the further comparative studies with the classical SRC methods demonstrate the superior fault recognition performance and low computation costs of SRC-SDD for rotating machinery fault diagnosis.

As a next step, we consider to incorporate the unsupervised or semi-supervised learning algorithm into the proposed SRC-SDD method for wide applications of rotating machinery fault diagnosis.

REFERENCES

[1] C. Peng, Z. Tang, W. Gui, Q. Chen, L. Zhang, X. Yuan, and X. Deng, “Review of key technologies and progress in industrial equipment health management,” IEEE Access, vol. 8, pp. 151764–151776, 2020.
[2] J. Guo, Z. Li, and M. Li, “A review on prognostics methods for engineering systems,” IEEE Trans. Rel., vol. 69, no. 3, pp. 1110–1129, Sep. 2020.
[3] L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan, and M. Wei, “A review on deep learning applications in prognostics and health management,” IEEE Access, vol. 7, pp. 162415–162438, 2019.
[4] D. Wang, K.-L. Tsui, and Q. Miao, “Prognostics and health management: A review of vibration based bearing and gear health indicators,” IEEE Access, vol. 6, pp. 665–676, 2018.
[5] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, “Machinery health prognostics: A systematic review from data acquisition to RUL prediction,” Mech. Syst. Signal Process., vol. 104, pp. 799–834, May 2018.
[6] T. Wang, Q. Han, F. Chu, and Z. Feng, “Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review,” Mech. Syst. Signal Process., vol. 126, pp. 662–685, Jul. 2019.
[7] J. Antoni and R. B. Randall, “The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines,” Mech. Syst. Signal Process., vol. 20, no. 2, pp. 308–331, Feb. 2006.

[8] G. L. McDonald and Q. Zhao, “Multipoint optimal minimum entropy deconvolution and convolution fix: Application to vibration fault detection,” Mech. Syst. Signal Process., vol. 82, pp. 461–477, Jun. 2017.

[9] X. Ye, Y. Hu, J. Shen, R. Feng, and G. Zhai, “An improved empirical mode decomposition based on adaptive weighted rational quartic spline for rolling bearing fault diagnosis,” IEEE Access, vol. 8, pp. 123813–123827, 2020.

[10] S. Lu, Q. He, T. Yuan, and F. Kong, “Online fault diagnosis of motor bearing via stochastic-resonance-based adaptive filter in an embedded system,” IEEE Trans. Syst., Man, Cybern. Syst., vol. 47, no. 7, pp. 1111–1122, Jul. 2017.

[11] Z. Feng, M. Liang, and F. Chu, “Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples,” Mech. Syst. Signal Process., vol. 38, no. 1, pp. 165–205, Jul. 2013.

[12] G. Yu, “A concentrated time–frequency analysis tool for bearing fault diagnosis,” IEEE Trans. Instrum. Meas., vol. 69, no. 2, pp. 371–381, Feb. 2020.

[13] Z. He, X. Tu, W. Bao, Y. Hu, and F. Li, “Second-order transient–extracting transform with application to time-frequency filtering,” IEEE Trans. Instrum. Meas., vol. 69, no. 8, pp. 5428–5437, Aug. 2020.

[14] R. Yan, R. X. Gao, and X. Chen, “Wavelets for fault diagnosis of rotary machines: A review with applications,” Signal Process., vol. 96, pp. 1–15, Mar. 2014.

[15] Y. Kong, T. Wang, and F. Chu, “Meshing frequency modulation assisted empirical wavelet transform for fault diagnosis of wind turbine planetary ring gear,” Renew. Energy, vol. 132, pp. 1373–1388, Mar. 2019.

[16] Y. Kong, T. Wang, and F. Chu, “Adaptive TQWT filter based feature extraction method and its application to detection of repetitive transients,” Sci. China Technol. Sci., vol. 61, no. 10, pp. 1556–1574, Oct. 2018.

[17] W. He, Y. Ding, Y. Zi, and I. W. Selesnick, “Sparsity-based algorithm for detecting faults in rotating machines,” Mech. Syst. Signal Process., vol. 72–73, pp. 66–82, May 2016.

[18] G. Cai, I. W. Selesnick, S. Wang, W. Dai, and Z. Zhu, “Sparsity-enhanced signal decomposition via generalized minimax-concave penalty for gearbox fault diagnosis,” J. Sound Vib., vol. 432, pp. 213–234, Oct. 2018.

[19] Z. Zhao, S. Wang, W. Xu, S. Wu, D. Wong, and X. Chen, “Sparsity-assisted fault feature enhancement: Algorithm-aware versus model-aware,” IEEE Trans. Instrum. Meas., vol. 69, no. 9, pp. 7004–7014, Sep. 2020.

[20] X. Dai and Z. Gao, “From model. signal to knowledge: A data-driven perspective of fault detection and diagnosis,” IEEE Trans. Ind. Informat., vol. 9, no. 4, pp. 2226–2238, Nov. 2013.

[21] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, “Deep learning and its applications to machine health monitoring,” Mech. Syst. Signal Process., vol. 115, pp. 213–237, Jan. 2019.

[22] J. Wright, Y. Li, L. Signal, Y. Zhang, “A new convolutional neural network-based data-driven fault diagnosis method,” IEEE Trans. Ind. Electron., vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[23] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, “Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing,” Mech. Syst. Signal Process., vol. 100, pp. 743–765, Feb. 2018.

[24] C. Sun, M. Ma, Z. Zhao, and X. Chen, “Sparse deep stacking network for fault diagnosis of motor,” IEEE Trans. Ind. Informat., vol. 14, no. 7, pp. 3261–3270, Jul. 2018.

[25] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, “Gearbox fault diagnosis using a deep learning model with limited data sample,” IEEE Trans. Ind. Informat., vol. 16, no. 10, pp. 6263–6271, Oct. 2020.

[26] Y. Bengio, “Learning deep architectures for AI,” Found. Trends Mach. Learn., vol. 2, no. 1, pp. 1–127, 2009.

[27] J. Wright, Y. Ma, I. Mairal, G. Sapiro, T. S. Huang, and S. Yan, “Sparse representation for computer vision and pattern recognition,” Proc. IEEE, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[28] X. Mei and H. Ling, “Robust visual tracking and vehicle classification via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 11, pp. 2259–2272, Nov. 2011.

[29] J. Wright, A. Y. Yang, A. Ganesh, S. Shankar Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.
TIANYANG WANG received the B.S. and Ph.D. degrees in mechanical engineering from Beijing Jiaotong University, Beijing, China, in 2008 and 2015, respectively. From 2012 to 2013, he was a Visiting Scholar with Ottawa University, Ottawa, Canada. From 2015 to 2018, he was a Postdoctoral Research Fellow with the Department of Mechanical Engineering, Tsinghua University, where he has been an Assistant Research Fellow, since 2018. His recent research interests include signal processing and fault diagnosis of rotating machineries.

ZHAOYE QIN received the B.S. and M.S. degrees in mechanical engineering from Northeastern University, Shenyang, China, in 2003 and 2006, respectively, and the Ph.D. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2010. From 2010 to 2013, he was a Postdoctoral Research Fellow with the Department of Mechanical Engineering, Tsinghua University. From 2016 to 2017, he was a Visiting Researcher with the University of Toronto, Toronto, Canada. Since 2017, he has been an Associate Professor with the Department of Mechanical Engineering, Tsinghua University. His research interests include rotating machinery dynamics and fault diagnosis, vibration control, composite materials, and vibration energy harvesting.

FULIE CHU received the B.S. degree in mechanical engineering from the Jiangxi University of Science and Technology, Ganzhou, China, in 1982, the M.S. degree in applied mechanics from Tianjin University, Tianjin, China, in 1985, and the Ph.D. degree in mechanical engineering from the University of Southampton, Southampton, U.K., in 1994. He is currently a Professor with the Department of Mechanical Engineering, Tsinghua University, Beijing, China. He has authored over 300 academic articles in the research areas of rotating machinery dynamics, machine fault diagnosis, nonlinear vibration, and vibration control.

Prof. Chu is the Vice President of the Chinese Society for Vibration Engineering (CSVE), the Executive Director of Fault Diagnosis Chapter in CSVE, and a member of the rotor dynamics Chapter in the International Federation for Promotion of Mechanism and Machine Science (IFToMM). He is a member of the editorial boards of over ten academic journals. He was a recipient of the Second Prize of National Natural Science Award in 2003, the National Science Fund for Distinguished Young Scholars in 2004, the Second Prize of Natural Science Award of Ministry of Education in 2006, the First Prize of Natural Science Award of Ministry of Education in 2011. From 2014 to 2019, he was selected as the most cited researchers for exceptional research performance in the field of Mechanical engineering by Elsevier.