

1639. Shift invariant sparse coding ensemble and its application in rolling bearing fault diagnosis

Xinqing Wang¹, Huijie Zhu², Ting Rui³, Yanfeng Li⁴, Tianshuai Liu⁵, Mengxi Liu⁶

^{1, 2, 3, 4, 5}College of Field Engineering, PLA University of Science and Technology, Nanjing, P. R. China

⁶Taicang State Taxation, Suzhou, P. R. China

²Corresponding author

E-mail: ¹rwuwaye@126.com, ²zhuhuijiehao@163.com, ³rtinguu@sohu.com, ⁴suoweiyiren163@163.com, ⁵1016002562@qq.com, ⁶lzmzxx1990@sina.com

(Received 4 December 2014; received in revised form 25 January 2015; accepted 10 February 2015)

Abstract. This paper proposes an automatic diagnostic scheme without manual feature extraction or signal pre-processing. It directly handles the original data from sensors and determines the condition of the rolling bearing. With proper application of the new technique of shift invariant sparse coding (SISC), it is much easier to recognize the fault. Yet, this SISC, though being a powerful machine learning algorithm to train and test the original signals, is quite demanding computationally. Therefore, this paper proposes a highly efficient SISC which has been proved by experiments to be capable of representing signals better and making converges faster. For better performance, the AdaBoost algorithm is also combined with SISC classifier. Validated by the fault diagnosis of bearings and compared with other methods, this algorithm has higher accuracy rate and is more robust to load as well as to certain variation of speed.

Keywords: shift invariant sparse coding, fault diagnosis, AdaBoost, dictionary learning.

1. Introduction

Traditional fault diagnosis procedure contains signal pre-processing, feature extraction, feature selection and pattern recognition [1]. Many approaches have been developed for different machines, and much success has also been achieved in machinery fault diagnosis [2]. However, due to the complicate steps and a large number of parameters to setting, it involves sufficient professional knowledge and experience. When the fault mechanism of the new equipment is unknown, or when the worker lacks experience, it will be a challenge to diagnose the rolling bearing efficiently. Therefore, a desirable diagnostic system should be able to automate the fault detection process without requiring any prior knowledge, so that it can be easily used even by those without experience [3]. In order to automate fault recognition and satisfy the demand for high accuracy, this paper designed a diagnostic scheme with the help of some improved machine learning algorithms.

When mechanical fault features are weak or relatively similar, it is hard to achieve condition monitor through traditional signal processing or pattern recognition techniques. Recently, sparse coding evolving from compress sensing has been a fundamental and effective technique in such fields as neurology, signal processing and pattern recognition [4, 5]. It represents signals as sparse linear combination of over complete bases, and achieves the simplest signal representation through the adaptive learning of the best bases and coefficients. This automatic learning algorithm does not require manual selection of features, and the features learned from original signals tend to represent the essence of signals [5]. Sparse coding is also a robust and adaptive pattern recognition algorithm [6]. Sparse coding for pattern recognition has been successfully applied in many fields such as face recognition and sound processing [7].

Due to the fact that the feature in the section of a signal may locate at any time, it is necessary to use many bases to represent the same feature for conventional sparse coding. The technique of shift invariant sparse coding (SISC) [8, 9] could represent such features sparsely, which describes a signal as the convolutions of bases and coefficients. Liu [10] is the first to apply SISC to rolling bearing fault diagnosis for feature extraction, and to demonstrate that its accuracy, robustness and adaptability are superior to traditional diagnostic methods. Tang [11] used SISC to find out and

extract, at the early failure stage, the latent fault features which are submerged or distorted by relatively strong background noise. Therefore, it can be reasoned that SISC is an excellent approach to feature extraction for fault detection.

However, some drawbacks still exist in practical application for SISC. Due to the convolution of bases and coefficients, SISC adds a dimension to the coefficient matrix compared to conventional sparse coding, and thus increases computation in solving the best coefficients and bases. Conventional approaches for solving SISC may easily be trapped into local optimal solution. In such case, an efficient SISC with better performance is necessary. Therefore, this paper proposes an improved shift invariant sparse coding (ISISC). In dictionary learning procedure, the optimum problem is simplified into a small linear equation, and thus can be solved directly by Cholesky decomposition. What's more, an improved orthogonal matching pursuit (OMP) and an online learning algorithm are also adopted for higher efficiency.

In practice of fault diagnosis, the normal and abnormal signals are very similar at the early failure stage. The strong background noise and variety of operating condition lead to more latent discriminating features, which requires that the performance of machine learning method for fault diagnosis be enhanced as much as possible, and that the classifiable ability of SISC be further improved. In order to improve the classifiable performance of conventional sparse coding, some researchers add discrimination constraint to the sparsity in dictionary learning process [7], so that the same classes have similar sparse codes and different classes have discriminating codes. But this method will make some fast dictionary learning algorithm unavailable and the optimization no longer convex. Ensemble learning is an accepted effective algorithm of machine learning [12], and AdaBoost algorithm [13] is a valid approach to ensemble learning. AdaBoost algorithm does not change the optimization function of ISISC, therefore, this paper introduces AdaBoost algorithm into ISISC for better pattern recognition.

2. ISISC ensemble

2.1. Introduction of SISC

In the theory of SISC [9], by using an over complete dictionary matrix $[\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_M] \in \mathbb{R}^{Q \times M}$ which contains M prototype bases for columns, a signal $\mathbf{s}_k \in \mathbb{R}^N$ can be represented as the sum of the convolutions of these bases and coefficients:

$$\mathbf{s}_k = \sum_{m=1}^M \mathbf{d}_m * \mathbf{x}_{m,k} + \boldsymbol{\varepsilon}, \tag{1}$$

where $*$ indicates convolution, coding coefficient $\mathbf{x}_{m,k} \in \mathbb{R}^P$ is sparse, $\boldsymbol{\varepsilon}$ is a normalized Gaussian noise, $Q < N$, $P < N$, and $Q + P - 1 = N$.

To make the optimization convex, the basis \mathbf{d}_m and coefficient $\mathbf{x}_{m,k}$ are optimized repeatedly. The cost function for evaluating the representation of the signal set containing K signals is as follows [14]:

$$\psi(\theta) = \min_{d,x} \sum_{k=1}^K \left\| \mathbf{s}_k - \sum_{m=1}^M \mathbf{d}_m * \mathbf{x}_{m,k} \right\|_2^2 + \beta \cdot \sum_{m,k} \|\mathbf{x}_{m,k}\|_1, \tag{2}$$

where $\|\cdot\|_2$ is l^2 norm, β is a parameter, which determines the degree of sparsity. The basis \mathbf{d}_m needs to be normalized, namely $\|\mathbf{d}_m\|_2^2 = 1$. If \mathbf{d}_m and $\mathbf{x}_{m,k}$ are solved at the same time, then the optimization problem is no-convex, making it more difficult to get a stable and convergent solution. As a result, the basis \mathbf{d}_m and coefficient $\mathbf{x}_{m,k}$ are optimized iteratively and repeatedly according to the maximum posterior probability estimation until convergence. Many experiments

have validated this approach as highly effective [4]. Many scholars have studied how to solve and simplify this problem. Smith [9] and Blumensath [8] proposed the most common optimization procedure. They directly updated the bases based on gradients, and successfully employed this way in voice and music processing. Grosse [14] and Tang [11] transformed signals into frequency domain, and used the gradients to solve the optimal bases. The gradient-based SISC (GSISC) method by Smith and Blumensath is as follows:

Let $\mathbf{e}_k = \mathbf{s}_k - \sum_{m=1}^M \mathbf{d}_m * \mathbf{x}_{m,k}$, the derivation of a given cost function $\psi(\theta)$ with respect to the q th element of basis \mathbf{d}_m is [8, 9]:

$$\frac{\partial \psi(\theta)}{\partial d_m^q} = - \sum_{n,k} e_k^n x_{m,k}^{n-q+1}, \tag{3}$$

where $1 \leq n \leq N, 1 \leq n - q + 1 \leq P, 1 \leq k \leq K$.

Similarly, the derivation of a given cost function $\psi(\theta)$ with respect to the p th element of coefficient vector $\mathbf{x}_{m,k}(p)$ is as [8, 9]:

$$\frac{\partial \psi(\theta)}{\partial x_{m,k}^p} = - \sum_q e_k^{q+p-1} d_m^q + \beta \cdot \text{sgn}(x_{m,k}^p), \tag{4}$$

where $1 \leq q + p - 1 \leq N, 1 \leq q \leq Q$.

After using the gradients in Eq. (3) and (4) to optimize the bases and coefficients iteratively until convergence is reached, the original signals can be sparsely represented by learned bases and coefficients. Besides, the coding coefficients could be also solved by OMP, FOCUSS algorithm, etc. As outlined above, the gradients used in these optimization algorithms are estimated through the numerical difference, and so the exact step length per iteration is hard to be taken, and the best evaluation in each iteration shall not guaranteed.

2.2. The procedure of ISISC

The efficient K-SVD algorithm [4] updates the bases one by one, rather than all at once. Inspired by this idea, the bases of ISISC are also updated one by one in the dictionary learning process. First, due to the sparsity of coefficients, the nonzero coefficients are extracted to markedly reduce the scale of the optimization problem. Then, the convolution problem is converted into a small scale symmetric positive linear equation, which can be directly solved by Cholesky algorithm without being influenced by former iterative result. So this scheme converges faster than GSISC. In the coefficient update stage, higher efficiency is achieved by the employment of OMP, fast convolution algorithm and extraction of only nonzero coefficients.

2.2.1. Dictionary update

In dictionary updating phase the coefficients are unchanged, and only the bases are updated. The cost function is simplified as:

$$\begin{aligned} \bar{\psi}(\theta) &= \min_{\mathbf{d}} \sum_{k=1}^K \left\| \mathbf{s}_k - \sum_{m=1}^M \mathbf{d}_m * \mathbf{x}_{m,k} \right\|_2^2 \\ &= \min_{\mathbf{d}} \sum_{k=1}^K \left\| \left(\mathbf{s}_k - \sum_{m \neq i}^M \mathbf{d}_m * \mathbf{x}_{m,k} \right) - \mathbf{d}_i * \mathbf{x}_{i,k} \right\|_2^2 = \min_{\mathbf{d}} \sum_{k=1}^K \left\| \mathbf{E}_{i,k} - \mathbf{d}_i * \mathbf{x}_{i,k} \right\|_2^2, \end{aligned} \tag{5}$$

where the matrix $\mathbf{E}_{i,k}$ stands for the error of all the bases except the i th basis with respect to the

k th signal.

Here is introduction on how to convert updating \mathbf{d}_i into solving a linear equation with regard to \mathbf{d}_i . Because $\mathbf{d}_i * \mathbf{x}_{i,k} = \mathbf{x}_{i,k} * \mathbf{d}_i$, the optimum problem with only the k th signal is analyzed first. The optimization of Eq. (5) equals to the solution of Eq. (6):

$$\begin{pmatrix} x_{i,k}^1 & & & & & \\ x_{i,k}^2 & x_{i,k}^1 & & & & \\ \vdots & x_{i,k}^2 & \ddots & & & \\ x_{i,k}^P & \vdots & \ddots & x_{i,k}^1 & & \\ & x_{i,k}^P & \ddots & x_{i,k}^2 & & \\ & & \ddots & \vdots & & \\ & & & x_{i,k}^P & & \end{pmatrix} \cdot \begin{pmatrix} d_m^1 \\ d_m^2 \\ \vdots \\ d_m^Q \end{pmatrix} = \begin{pmatrix} E_{i,k}^1 \\ E_{i,k}^2 \\ \vdots \\ E_{i,k}^N \end{pmatrix}. \quad (6)$$

If the left matrix of Eq. (6) is taken as a special Toeplitz matrix with regard to coefficient $\mathbf{x}_{i,k}$, then Eq. (6) can be rewritten as $Toep(\mathbf{x}_{i,k}) \cdot \mathbf{d}_i = \mathbf{E}_{i,k}$. In view of the sparsity of coefficient $\mathbf{x}_{i,k}$, many rows in the matrix $Toep(\mathbf{x}_{i,k})$ are all zeros and these zero rows have no effect on the update of bases. Delete these zero rows from matrix $Toep(\mathbf{x}_{i,k})$ and those corresponding rows in matrix \mathbf{E}_i , then Eq. (6) can be rewritten as $Toep(\tilde{\mathbf{x}}_{i,k}) \cdot \mathbf{d}_i = \tilde{\mathbf{E}}_{i,k}$. When all the K signals are taken consideration, the optimization of the cost function is represented as:

$$\begin{pmatrix} Toep(\tilde{\mathbf{x}}_{i,1}) \\ Toep(\tilde{\mathbf{x}}_{i,2}) \\ \vdots \\ Toep(\tilde{\mathbf{x}}_{i,K}) \end{pmatrix} \cdot \mathbf{d}_i = \begin{pmatrix} \tilde{\mathbf{E}}_{i,1} \\ \tilde{\mathbf{E}}_{i,2} \\ \vdots \\ \tilde{\mathbf{E}}_{i,K} \end{pmatrix}. \quad (7)$$

When Eq. (7) is abbreviated as $\mathbf{X} \cdot \mathbf{d}_i = \mathbf{E}$, and based on the least square method, the updated basis can be obtained as $\mathbf{d}_i = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{E})$. Because matrix $\mathbf{X}^T \mathbf{X} \in \mathbb{R}^{Q \times Q}$, and $Q \ll N$ in general, the above equation is transformed into a small scale of linear equation, which can be directly solved by Cholesky factorization, or LU factorization for the optimal value. In theory, during the process of dictionary learning, the update relation of basis \mathbf{d}_i given in Eq. (7) is the best that can be achieved provided that coefficients and other bases are fixed, and the solution will not be affected by the former iterative result. But since it is hard to choose the step length accurately, GSISC converged relatively slower, and the update is affected by the former iterative result.

All the bases are randomly updated one by one so as to get the best solution of the basis in this round. Afterwards each basis is normalized as $\mathbf{d}_i = \mathbf{d}_i / \|\mathbf{d}_i\|_2^2$.

2.2.2. Shift invariant OMP

OMP is chosen to solve the coding coefficients, and in this process all the bases are fixed, so that only the coefficients are optimized. As a new version of matching pursuit, OMP is also an iterative greedy algorithm [15]. After choosing the best suitable basis in each iteration, OMP would make all the bases to be orthogonal to each other and project them on the original signal again.

The u points translation of the basis \mathbf{d}_i is padded with zeros so that the new basis is as long as original signal, which is represented as $\mathbf{g}_{i,u}$, and $\|\mathbf{g}_{i,u}\| = 1$. $\{\mathbf{g}_{i,u} | 1 \leq i \leq Q, 1 \leq u \leq P\}$ is the over complete dictionary for sparse representation. The procedure of shift invariant OMP is as follows [15]:

- (1) Set \mathbf{s}_k as the signal for representation, initialize the index set $\Psi_0 = \emptyset$, the iteration counter

$n = 1$, and the residual $\mathbf{r}^1 = \mathbf{s}_k$.

(2) Find the index of the basis that satisfies:

$$|\langle \mathbf{r}^n, \mathbf{g}_{i,u}^n \rangle| = \sup_{1 \leq i \leq Q} \left(\sup_{1 \leq u \leq P} |\langle \mathbf{r}^n, \mathbf{g}_{i,u}^n \rangle| \right). \quad (8)$$

The fast convolution can be applied for fast calculation when calculating $\sup_{1 \leq u \leq P} |\langle \mathbf{r}^n, \mathbf{g}_{i,u}^n \rangle|$.

(3) Augment the matrix of chosen bases: $\Psi_n = \Psi_{n-1} \cup \{\mathbf{g}_{i,u}^n\}$.

(4) Because of the sparsity of the chosen bases $\mathbf{g}_{i,u}^n$ (namely column vectors), the matrix Ψ_n has many rows composed of zeros. The nonzero rows of matrix $\tilde{\Psi}_n$ are extracted from original matrix Ψ_n , and the corresponding elements of signal \mathbf{s}_k are extracted as the vector $\tilde{\mathbf{s}}_k$. This technique will significantly reduce the size of the matrix and enhance the efficiency. According to the least square method, the coefficients are estimated: $\mathbf{a}_n = (\tilde{\Psi}_n^T \tilde{\Psi}_n)^{-1} \tilde{\Psi}_n^T \tilde{\mathbf{s}}_k$, and the residual is $\mathbf{r}^n = \mathbf{s}_k - \mathbf{a}_n \Psi_n$.

(5) Increment $n = n + 1$, and return to Step 2 if n is smaller than the maximum iterations. Otherwise output the estimated reconstructed signal $\hat{\mathbf{s}}_k = \mathbf{a}_n \Psi_n$ and its residual \mathbf{r}^n .

In addition, the procedure can also be accelerated by batch-OMP [16] for more efficiency. Worked with proposed OMP, the sparse coding will be evaluated at an extremely faster speed.

2.2.3. Online learning

Mairal [17] is the first to introduce online learning to sparse coding to increase the computational speed. Tang [11] applied this method to SISC. In Tang's algorithm, the dictionary is updated only by one sample in each iteration, causing the dictionary to match only the current sample well and easily leading to local optimal solution. This paper proposes a new online learning algorithm to overcome this shortcoming.

Step 1, divide the training sample set into m small sub sets, and the m th sub set is O_m . The iteration number $n = 0$, the candidate training set Ψ_0 is null set. Initiate the dictionary randomly.

Step 2, let $n = n + 1$, and $\Psi_n = \{\Psi_{n-1}, O_n\}$.

Step 3, update the coefficients by shift invariant OMP of all the samples in Ψ_n .

Step 4, update the dictionary by ISISC, using all the samples in Ψ_n .

Step 5, if $n = m$, go to step 6, otherwise go to step 2.

Step 6, let $n = n + 1$, using all the samples to update the coefficients with shift invariant OMP, and update the dictionary by ISISC.

Step 7, if n equals the max iteration number, output the trained dictionary, otherwise go to step 6.

While applying the algorithm above, it is necessary to just use few samples to update the dictionary for faster convergence, because the dictionary is far from the best dictionary in the former phase. However, in the later phase, if only few samples are used, it may get trapped in local optimal solution, hence much more samples are applied for fine tuning.

2.3. SISC ensemble for pattern recognition

Sparse coding and pattern recognition have natural relationship [6]. Under the limitation of sparsity, the most closely related basis with the original signal will be automatically selected, and the other bases are refused. The robustness and accuracy of sparse coding for pattern recognition have been validated in the field of face recognition and voice processing. The original dictionary can be constructed directly by selecting some training samples, and can be learned through dictionary learning algorithm. Dictionary learning [5] is able to learn the internal structures and main features of original signals through a group of sparse bases, and it is insensitive to Gaussian noise. In this paper, the ISISC proposed in Section 2.2 is used for dictionary learning, and one sub

dictionary per class is learned.

In this investigation, the sub dictionaries are not combined together in order to avoid destroying the structure of each sub dictionary. But under the same sparsity of T_0 , each sub dictionary is employed respectively for sparse representation. When the label of the sample and the label of the sub dictionary are the same, the reconstruction error should be the lowest, which means that the label of the dictionary with lowest reconstruction error is expected to be the label of the sample.

Unfortunately, real-world vibration signals are often contaminated by a lot of noise, and the number of available training sample is limited. Thus it is necessary to further improve the classification ability of sparse coding. Some scholars added distinctive constraints to sparsity when training dictionary so that the dictionaries can be more discerning. However, this approach will increase the difficulty of optimization and make some fast algorithm potentially unable to work well, or even make convergence unstable. Ensemble learning introduced in recent years is a very powerful machine learning method [12]. It synthesizes a few different classifiers, and can get better performance than the individual classifier that composes it. AdaBoost algorithm [13], proposed by Freund and Schapire, is a valid approach to ensemble learning. The same group of training samples are used, but the weights of samples are adaptively changed after a classifier is trained. Therefore, the error sample will be paid more attention by the next classifier. Lastly the combination of classifiers, whose individual decisions are weighted by voting, will exhibit better performance. It has been verified that when the error of individual classifier is lower than 50 %, the error of combination will descend exponentially close to zero [13].

Although the sparse coding is not the same as conventional classifiers, the increase in the weight of an error sample will enhance its ability to recognize the pattern of a sample. The rise of a sample's weight will result in the fact that the structure of the sub-dictionary learned is closer to the error sample. That is to say, the dictionary has superior ability to represent this sample, while keeping other sub-dictionaries representative ability unchanged. Therefore the reconstruction error of the sample by the new weighted sub-dictionary is relatively decreased, and the probability of being categorized as required is increased. The algorithm of ISISC ensemble is presented in Table 1 [18].

Table 1. Proposed algorithm of ISISC ensemble

<p>AdaBoost (S, ISISC, T) Input: \mathbf{S}, training sample set. Each class has K training samples $\{\mathbf{s}_1^l, \dots, \mathbf{s}_k^l\}$ with labels $l \in \{1, 2, \dots, L\}$. ISISC, the base-learning algorithm. T, the number of ISISC ensemble. Initialize: the weight of sample \mathbf{s}_k^l as $w_k^l = 1/KL$, $t = 0$, $\varepsilon_0 = 0$. Do while $t \leq T$ and $\varepsilon_t < 0.5$. 1. Normalize the weights: $\sum_{k=1, l=1}^{K, L} w_k^l(t) = 1$. 2. Train the tth ISISC classifier with adaptive updated weights, and recognize all the training samples. If the sample is right, namely $f_t(\mathbf{s}_k^l) = l$, then $e_k^l = 0$, otherwise $e_k^l = 1$. 3. Calculate $\varepsilon_t = \sum_{k=1, l=1}^{K, L} w_k^l(t)e_k^l(t)$, and let $\alpha^t = 0.5 \log[(1 - \varepsilon_t)/\varepsilon_t]$. 4. Update the weight as: $w_k^l(t + 1) = w_k^l(t) \times \begin{cases} \exp(-\alpha^t), & f_t(\mathbf{s}_k^l) = l, \\ \exp(\alpha^t), & f_t(\mathbf{s}_k^l) \neq l. \end{cases}$ 5. Increment the count number: $t = t + 1$. Output: the synthesis of all the classifiers decisions: $f(s_k) = \operatorname{argmax}_{1 \leq l \leq L} \sum_{j=1}^t \alpha(j) f_j(s_k) = l$, where x is true $x = 1$, otherwise $x = 0$.</p>

This paper fused the AdaBoost algorithm into ISISC to enhance the successful rate for recognition faults, and the diagnostic procedure is shown in Fig. 1. The signals under different loads have diverse energies. By normalizing them, it is able to guarantee that the relative weights of samples are identical. The sub-dictionary is trained by ISISC, one for each class, in sequence

according to the labels. After the first classifier was trained, and the weights of all training samples are then updated according to the method demonstrated in Table 1. Subsequently the sub-dictionaries of the next classifier are trained. Following this way, all the ISISC classifiers are to be trained. For an unlabeled sample, firstly, it is to be reconstructed under sparsity T_0 respectively by using each sub dictionary of the first classifier; secondly, it is to be classified based on the minimum reconstruction error. In such way, the sample is identified by all the ISISC classifiers. Eventually all the decisions are synthesized for the final judgment in the light of the weighted algorithm in Table 1.

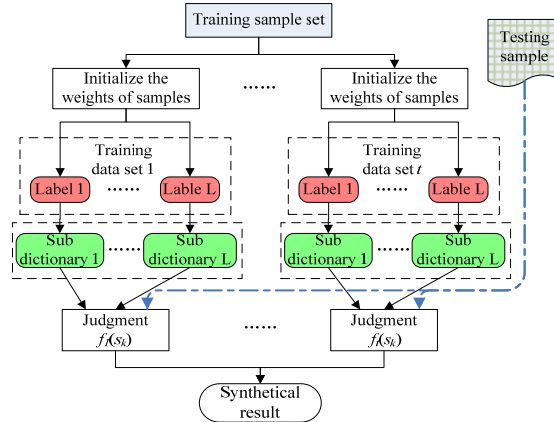


Fig. 1. ISISC ensemble procedure

3. Experiment analysis and discussion

3.1. Comparison between ISISC and GSISC

A simulated pulse sequence of vibration signal caused by a single point defect on rolling bearing is as follows:

$$s = \sum_{m=1}^M 20 \cdot (\sin(60\pi t) + 2)y(t - mT - \tau_i) + noise, \quad (9)$$

where t is time, $y(t)$ has a damped sinusoids function and its amplitude is modulated by a sinusoids function. i is the count number of pulses, τ_i is the beginning time of pulse. Here τ_i is set as a random number in order to test the adaptability of SISC. T is the period of the pulses, and here it equals 120 points. The attached noise is Gaussian noise, and its signal to noise ratio (SNR) is -2 dB. $y(t) = e^{pt} \sin(2\pi f_b t)$ is a pulse resonance induced by the single point defect, and it is normalized in this study. $p = 1200$ is the damping coefficient, and f is oscillation frequency, which indicates the damped natural frequency of the system.

The signal has 10240 points, and it is averagely separated into ten samples. GSISC and ISISC are respectively used to represent the signal set for comparison. The length of a basis is set as 60 points, and only one basis is learned. For GSISC optimization, the conjugate gradient method is chosen, and the step length is determined by both the golden section search and parabolic interpolation. The sparsity is set as 9 in ISISC, and the online learning algorithm is not adopted in this experiment.

The convergence curves of residual energy is shown in Fig. 2. The original basis, the bases learned by ISISC and GSISC are illustrated in Fig. 3. ISISC converges to the optimal solution within a few iterations, and its learned basis resembles the original pulse resonance. Yet, GSISC converges considerably slowly, and its solution is slightly worse than that of ISISC. Moreover,

there is a little distortion with its learned basis.

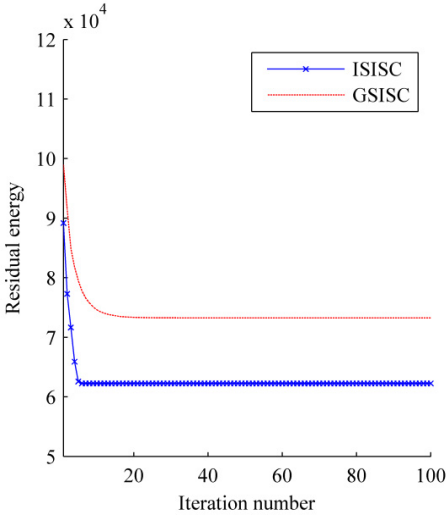


Fig. 2. Residual energy

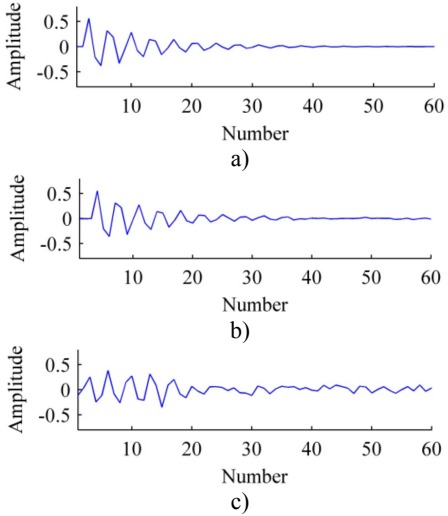


Fig. 3. Bases: a) origin, b) ISISC, c) GSISC

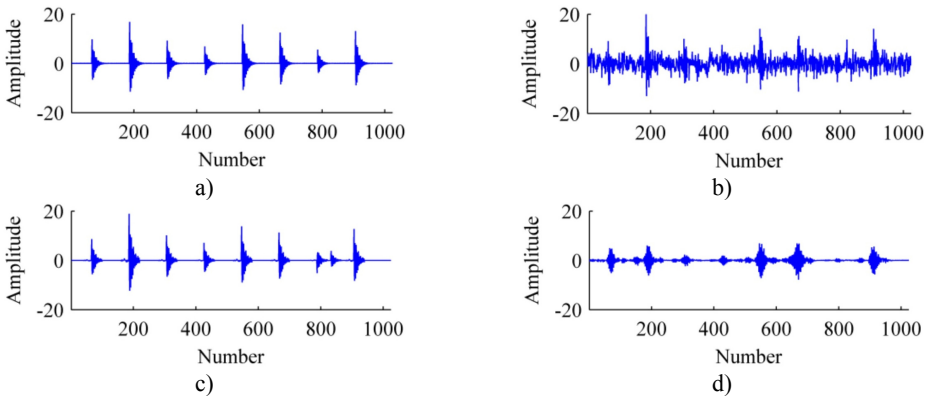


Fig. 4. The signal: a) origin, b) added with noises, c) ISISC, d) GSISC

Fig. 4 displays an original sample, and corresponding reconstructions by ISISC, GSISC. The signal is seriously contaminated by noises, but ISISC still excellently recovers it. Unlike ISISC, the reconstruction of GSISC distorts and contains a few noises, and the peaks of pulses are not well preserved.

To comprehensively appraise the performances of these two algorithms, the original signals are mounted with noises of different SNR values. The SNR and MSE values of the 100th iteration are applied for evaluation. Fig. 5 shows the average results of ten times. No matter how many noises added to the original signals, ISISC always outperforms GSISC in both SNR and MSE. In contrast to GSISC and in consideration of the new dictionary updated algorithm and the experiment, ISISC is more efficient, exhibits faster convergence, is not as easy to fall into local optima, and has higher probability to find better optimal solution.

3.2. Application of ISISC ensemble to mechanical fault diagnosis

Rolling bearing is one of the most widely used elements in machinery, and its status is vital to the whole system. The database used in this paper is obtained from the Case Western Reserve University Bearing Data Center [19], which is considered as a benchmark for machinery diagnosis.

Single point faults with diameters of 0.007, 0.014, 0.021 and 0.028 inches (1 inch = 2.54 cm) are introduced separately at the outer race (at 6:00 directions), inner race and ball. Tests are carried out respectively under loads of 0, 1, 2, 3 hp (1 hp = 746 W), with corresponding shaft speed being 1797, 1772, 1750, 1730 rpm. The sampling frequency is 12 kHz. In order to determine the location and the severity of the fault, the data are divided into 10 classes, and each class contains four data with the loads of 0, 1, 2, 3 hp. The relationship between labels and data is shown in Table 2. Normal indicates the bearing is Normal. IR, B and OR respectively represent the inner race fault, ball fault and outer race fault. The number followed suggests the fault severity. For example, OR007 represents the outer race of the bearing is fault, and the fault diameter is 0.007 inch.

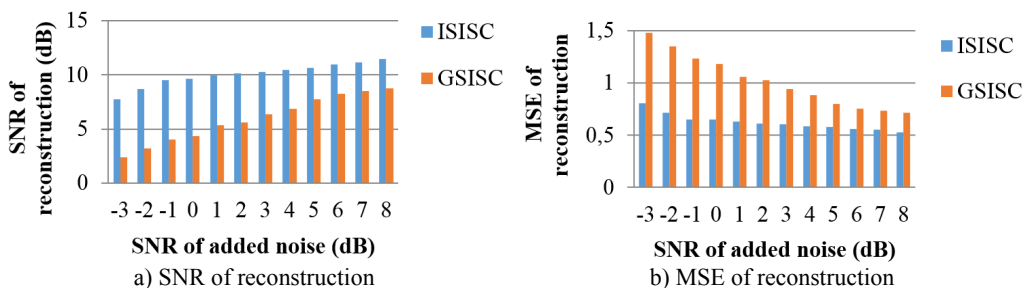


Fig. 5. The performances of ISISC and GSISC with different noises

Table 2. The relationship between data and labels

Data	Normal	IR007	B007	OR007	IR014	B014	OR014	IR021	OR021	IR028
Label	1	2	3	4	5	6	7	8	9	10

Every data is truncated into time-series with a 2048-point window block. About 118 samples are generated from each data of normal condition, while each data under faulty condition contains approximately 59 samples. The ISISC ensemble proposed in this report is used to train and test the data. The number of classifiers in ensemble is five; each sub dictionary for one class T_0 contains 15 bases, and the sparsity is 26. The length of a basis is 150 points, and as its length will contain about 2 pulses in one basis, the basis will include not only the underlying information of resonance but also the fault feature frequency of each condition.

Liu [10] has applied this data set, employed SISC for feature extraction, and designed a linear classifier for identification. In his experiment, all the samples are smoothed for pre-processing. For the purpose of confirming the advantage of ISISC ensemble, two similar experiments are conducted. In the first experiment, 20 samples per data under every load are randomly selected as training set, and all the samples in the database are tested. This experiment is implemented by ISISC ensemble. The single ISISC classifier is also tested, only using 25 samples per each data for training. In Liu's experiment [10], half samples in each data are trained which is more than either ISISC ensemble or ISISC classifier. The average accuracy rate of 0-3 hp for each class is shown in Fig. 6.

Although ISISC ensemble uses the least training samples, its success rates are highest and stable in general. This demonstrates that ISISC ensemble can accurately determine the condition of bearing as well as fault severity. Moreover, ISISC classifier and ISISC ensemble both perform better than Liu's method on the whole. Apart from that, the efficiency has been markedly improved. Using a laptop with a dual-core 2.2 GHz CPU and DDR 667 2G memory, the training time of one class using Liu's method is approximately one hour, whereas the time of ISISC ensemble is about ten minutes and the time of ISISC classifier is as few as 2 minutes. In comparison with Liu's method of adding a classifier for recognition and pre-processing the samples, all the samples can be directly put into ISISC ensemble for fault recognition without any people's operation. Dong [20] employed principal components analysis to extract features which were input into Morlet kernel support vector machine for classification. He recognized the same

data of Normal, IR007, IR014, IR021 and IR028, with the accuracy rates are all as high as 100 % under the best parameters. By contrast, the accuracy rates of these data using ISISC ensemble are also 100 %, but the classes of ISISC ensemble are 10 which is more than that of Dong's experiment.

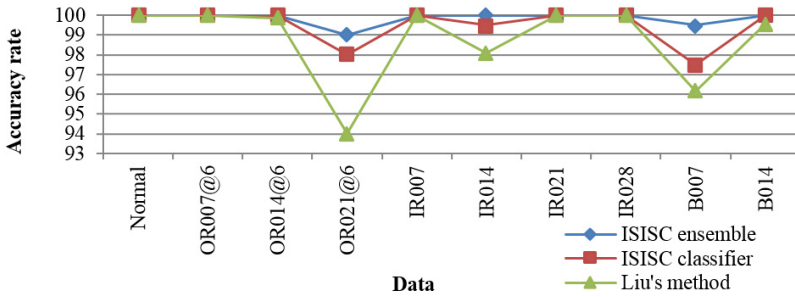


Fig. 6. The average accuracy rate of 0-3 hp for each class

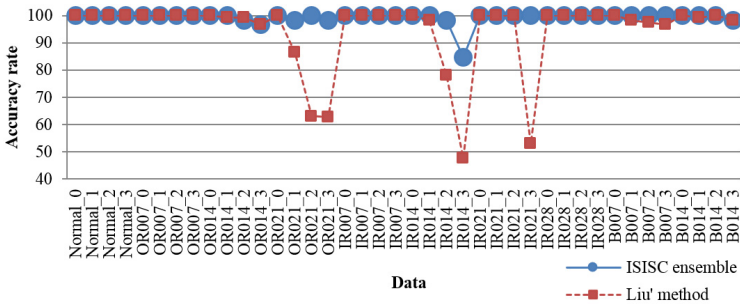


Fig. 7. The accuracy rates of each data under 0-3 hp

When the rolling bearing is running, its load frequently varies. The rolling bearing can hardly run at an accurate speed, and the speed usually varies in a small range. Therefore, the monitoring system needs to be robust enough against the alteration of load and the variation of speed within a small range. So another experiment is done, with only the data under 0 hp selected for training and the corresponding speed being 1797 rpm. In the training set, 50 samples per class are randomly chosen. All the data are selected for testing, namely all the data under 0, 1, 2, 3 hp, and corresponding speeds are 1797, 1772, 1750, 1730 rpm respectively. Liu [10] has also done the similar experiment, and the results of our experiment together with those of Liu's are listed in detail in Fig. 7. The accuracy rates of ISISC ensemble are higher than Liu's for all the data. The worst one of our method is above 80 %, while Liu's worst rate is below 50 %, so the stability of ISISC ensemble is stronger. Zhu [2] has applied K-SVD based on envelope spectrum to do similar experiment, but he's algorithm need to re-sample the data, so it was not robust enough to speed.

As mentioned above, it is evident that ISISC ensemble is more accurate than other methods. Apart from that, the most outstanding advantage of ISISC ensemble is that it lessens people's participation and is much easy for application. The signal pre-processing, feature extraction and pattern recognize are not needed in this algorithm, and all the learning processes are implemented by computer itself. ISISC ensemble can still make precise judgment even when the load varies or the speed changes in a certain range.

4. Conclusions

This paper presents an intelligent diagnostic scheme based on improved SISC algorithm, and AdaBoost algorithm. This ISISC ensemble directly analyzes the original signal measured from sensors, and does not involve such technique as pre-processing, feature extraction, feature

selection and so on. This diagnosis scheme is simple and demands less people's manipulation, thus avoiding subjective operation. In order to improve the efficiency and performance of conventional SISC algorithm, the proposed ISISC transforms the complicated dictionary update into solving a small linear equation. In addition, a shift invariant OMP and an online learning algorithm are further applied for accelerating convergence. Tested by simulated signals, ISISC has faster convergence speed and can achieve better solution than GSISC.

Finally, ISISC ensemble is applied to the diagnosis of bearings, and a comparison with other researches is attached. The experiments confirm that ISISC ensemble performs better than a single ISISC classifier. Compared to other approaches, ISISC ensemble improves the accuracy rate, and can also precisely determine the fault location as well as fault severity of bearings. In addition, it is robust to the load or the speed variation in a certain range. In conclusion, ISISC ensemble facilitates the diagnostic task, and has certain value for rolling bearing diagnosis.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 61472444, 61472392.

References

- [1] **Hong H., Liang M.** Fault severity assessment for rolling element bearings using the Lempel-Ziv complexity and continuous wavelet transform. *Journal of Sound and Vibration*, Vol. 320, Issue 1, 2009, p. 452-468.
- [2] **Zhu H., Wang X., Zhao Y., et al.** Sparse representation based on adaptive multiscale features for robust machinery fault diagnosis. *Proceedings of the Institution of Mechanical Engineers, Part C, Journal of Mechanical Engineering Science*, 2014.
- [3] **Kankar P., Sharma S. C., Harsha S.** Fault diagnosis of ball bearings using machine learning methods. *Expert Systems with Applications*, Vol. 38, Issue 3, 2011, p. 1876-1886.
- [4] **Aharon M., Elad M., Bruckstein A.** K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, Vol. 54, Issue 11, 2006, p. 4311-4322.
- [5] **Olshausen B. A.** Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, Vol. 381, Issue 6583, 1996, p. 607-609.
- [6] **Wright J., Yang A. Y., Ganesh A., et al.** Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, Issue 2, 2009, p. 210-227.
- [7] **Jiang Z., Lin Z., Davis L. S.** Learning a discriminative dictionary for sparse coding via label consistent K-SVD. *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, p. 1697-1704.
- [8] **Blumensath T., Davies M.** Sparse and shift-invariant representations of music. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, Issue 1, 2006, p. 50-57.
- [9] **Smith E. C., Lewicki M. S.** Efficient auditory coding. *Nature*, Vol. 439, Issue 7079, 2006, p. 978-982.
- [10] **Liu H., Liu C., Huang Y.** Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mechanical Systems and Signal Processing*, Vol. 25, Issue 2, 2011, p. 558-574.
- [11] **Tang H., Chen J., Dong G.** Sparse representation based latent components analysis for machinery weak fault detection. *Mechanical Systems and Signal Processing*, Vol. 46, Issue 2, 2014, p. 373-388.
- [12] **Polikar R.** Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, Vol. 6, Issue 3, 2006, p. 21-45.
- [13] **Freund Y., Schapire R. E.** A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, Vol. 55, Issue 1, 1997, p. 119-139.
- [14] **Grosse R., Raina R., Kwong H., et al.** Shift-invariance sparse coding for audio classification. *Conference on Uncertainty in AI*, 2007.
- [15] **Pati Y. C., Rezaifar R., Krishnaprasad P.** Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *IEEE The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, Vol. 1, 1993, p. 40-44.
- [16] **Rubinstein R., Zibulevsky M., Elad M.** Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *CS Technion*, Vol. 40, Issue 8, 2008, p. 40.

- [17] **Mairal J., Bach F., Ponce J., et al.** Online learning for matrix factorization and sparse coding. The Journal of Machine Learning Research, Vol. 11, 2010, p. 19-60.
- [18] **Hu Q., He Z., Zhang Z., et al.** Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. Mechanical Systems and Signal Processing, Vol. 21, Issue 2, 2007, p. 688-705.
- [19] **Loparo K. A.** Bearing vibration dataset. Case Western Reserve University, <http://www.eecs.cwru.edu/laboratory/bearing/download.htm>, 2003.
- [20] **Dong S., Sun D., Tang B., et al.** A fault diagnosis method for rotating machinery based on PCA and morlet kernel SVM. Mathematical Problems in Engineering, Vol. 2014, 2014, p. 8.



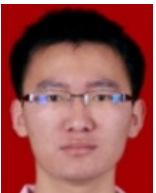
Xinqing Wang received Ph.D. degree in Mechanical Engineering from Tianjin University, Tianjin City, China, in 1998. Now he works as a Professor at PLA University of Science and Technology, China. His current research interests include mechanical-electronic engineering, fault diagnosis and signal processing.



Huijie Zhu received his B.S. degree in Automobile Service Engineering from Wuhan University of Technology, Wuhan City, China, in 2009. Now he is a Ph.D. candidate at PLA University of Science and Technology, China. His current research interests include signal processing, machine learning and machinery fault diagnosis.



Ting Rui received Ph.D. degree in Mechanical Engineering from Nanjing University of Aeronautics and Astronautics, Nanjing City, China, in 2001. Now he works as a Professor at PLA University of Science and Technology, China. His current research interests include computer vision, machine learning, multimedia, and video surveillance.



Yanfeng Li received his Master degree in Mechanical Engineering from PLA University of Science and Technology, Nanjing City, China, in 2012. Now he is a Ph.D. candidate. His current research interests include deep learning and mechanical system fault diagnosis.



Tianshuai Liu received his Bachelor degree in Mechanical Engineering from Wuhan University, Wuhan City, China, in 2012. Now he is a Master candidate. His current research interests include mechanical-electronic engineering and mechanical system fault diagnosis.



Mengxi Liu received her B.S. and M.S. degree in Accounting from Nanjing Agricultural University, Nanjing City, China, in 2012 and 2014, respectively. Now she is a researcher in Taicang State Taxation, Suzhou City, China. Her current research interests include data mining and statistics.