Deep Stack Dictionary learning for Fault Diagnosis of Rotating Machinery

Jinyang Jiao[1], Ming Zhao[2], and Jing Lin[3]

[1,2]*Shannxi Key Laboratory of Mechanical Product Quality Assurance and Diagnostic School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China*

jjy2015@stu.xjtu.edu.cn
zhaomingxjtu@mail.xjtu.edu.cn

[3]*State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710054, People's Republic of China*

jinglin@mail.xjtu.edu.cn

ABSTRACT

Effective feature extraction of rotating machinery has been a hot topic in the prognosis and health management. However, it is a challenging problem to extract periodic impulses under heavy background noise and other interference. In the last decade, deep learning and dictionary learning have been promising methods to extract feature information, which have made great achievement in the field of image, video denoising, etc. In this paper, via fusing the deep learning with dictionary learning, an algorithm called deep stack dictionary learning is proposed. This algorithm is trained in a layer-wise greedy manner so as to suppress the noise and highlight periodic impulses.

1. INTRODUCTION

Due to heavy background noise and random interference, it is a challenging task to develop effective signal processing techniques that extract fault characteristic from the measured vibration signals validly (Lei, Lin, He and Zi, 2011). Different from traditional representation algorithms, dictionary learning and deep neural network have received a lot of interest recently. Inspired by them, in this paper, we propose deep stack dictionary learning (DSDL) to extract the weak fault feature from the original vibration signals. The proposed algorithm mainly has two merits: 1). Dictionary learning can adaptively learn fault feature from the original signal without any prior knowledge. 2). A deep greedy layer-wise stack (GLWS) strategy (Bengio, Lamblin, Popovici and Larochelle, 2007) further improves the feature of learned periodic impulses, and extracts weak fault feature in the early stage. The effectiveness and robustness of the new method are validated by simulated signals and the vibration data measured from bearing test rigs.

The rest of the paper is organized as follows: the theory of K-SVD and GLWS is briefly introduced in Section 2. Then the proposed deep stack dictionary learning algorithm is illustrated in detail in Section 3. In Section 4, the effectiveness of the proposed method is validated using numerical simulation and datasets from rolling element bearings. Finally, conclusions are drawn in Section 5.

2. THEORETICAL BACKGROUND

In this section, we briefly introduce the theory of K-SVD and greedy layer-wise stack training, which are the theoretical basis of the proposed algorithm.

2.1. K-SVD

K-SVD, as one of the most popular dictionary learning algorithms, is proposed by Aharon, Elad and Bruckstein, (2006). Given training signals $\mathbf{y}$, it can be considered as solving the follow optimization problem iteratively:

$$\min_{\mathbf{D},\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_F^2 \le \varepsilon \text{ subjuct to } \forall i, \|x_i\|_0 \le T_0 \quad (1)$$

In the first stage, one can fix dictionary $\mathbf{D}$ and aim to find the best coefficient matrix $\mathbf{x}$ by decoupled to N distinct problems, which can be described as follows:

$$\min_{x_i} \|y_i - \mathbf{D}x_i\|_F^2 \le \varepsilon \text{ subjuct to } \forall i, \|x_i\|_0 \le T_0, \quad \text{for } i = 1, 2, ..., N. \quad (2)$$

In order to solve Eq.(2), greedy pursuit algorithms like orthogonal matching pursuit (OMP) (Tropp & Gilbert, 2007) can be used for finding a sub-optimal solution. In the second stage, one can search for a better dictionary given the representation coefficient, the penalty term can be rewritten as

$$\begin{aligned}\|\mathbf{Y}-\mathbf{DX}\|_F^2 &= \left\|(\mathbf{Y}-\sum_{j\neq k} d_j x_T^j)-d_k x_T^k\right\|_F^2 \\ &= \left\|E_k - d_k x_T^k\right\|_F^2\end{aligned} \tag{3}$$

The SVD or other approximate optimization method can be used to update **D** column by column. A complete description of K-SVD theory could be found in (Aharon at al., 2006; Rubinstein, Zibulevsky and Elad, 2008).

2.2. Greedy layer-wise stacked training

The strong representation ability of deep neural networks is attributed to its stacked hidden layers and GLWS training method. The GLWS was firstly introduced (Hinton & Salakhutdinov, 2006) to train a deep belief network (DBN) in 2006, whose basic idea is that after training the structure of current layer, the output can be used as the input to train the basic structure of subsequent layer. The basic structure of DBN and deep stacked autoencoder are restricted Boltzmann machine and autoencoder, respectively. We illustrate the idea of this method through deep stacked autoencoder in this paper.

An autoencoder (Vincent, Larochelle, Bengio and Manzagol, 2008) is a symmetrical neural network, which consists of two phases including encoder and decoder. The basic structure is shown in Figure 1, where encoder takes an input **x** and transforms it to a hidden representation **h** via a non-linear mapping as follow:

$$\mathbf{h} = f_\theta(\mathbf{Wx}+\mathbf{b}) \tag{4}$$

where f_θ is a non-linear activation function, **h** is the encoder vector obtained from **x**. Then, decoder network maps **h** back to the inputs in a similar way as follows:

$$\mathbf{x'} = g_{\theta'}(\mathbf{W'h}+\mathbf{b'}) \tag{5}$$

Model parameters including θ= [W, b, W', b'] are optimized to minimize the reconstruction error between **x**' and **x**.

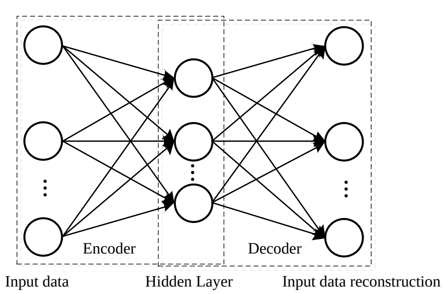

Figure 1. The structure of an autoencoder

The stacked autoencoder (SAE) is a deep neural network consisting of multiple layers of basis autoencoder (see Figure 2), in which the algorithm of greedy layer-wise training plays an important role. The outputs of each layer are wired to the inputs of each successive layer. In addition, the greedy layer-wise training method improves the training process so that the network hardly falls into local optimum solution.

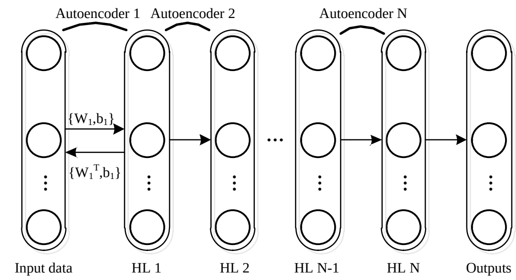

Figure 2. The structure of stack autoencoder (HL is short for Hidden layer)

3. DEEP STACK DICTIONARY LEARNING

Figure 3(a) shows the schematic diagram for dictionary learning, **Y** is the data, **D** is the dictionary and **X** is the feature of **Y** in **D**. Since a single level of dictionary learning yields a shallow feature representation of data, the concept of deep dictionary learning (DDL) has been recently proposed (Tariyal, Majumdar, Singh and Vatsa, 2016), which aims to learn deeper latent representations, see Figure 3(b). The idea of DLL is to learn multiple levels of dictionaries in a greedy fashion, in which the features from first layer (X_1) can be used as input to the second layer, and so on. Mathematical expressions at the second layer can be written as:

$$Y = D_1\varphi(D_2X_2) \tag{6}$$

where φ is activation function, it can be linear or non-linear.

Extending this idea, a multi-level DLL problem can be expressed as：

$$Y = D_1\varphi(D_2\varphi(...\varphi(D_NX_N))) \tag{7}$$

Because of space limitations, a detailed description and optimal solution of DDL could be found in (Tariyal et al., 2016; Singhal & Majumdar, 2017).

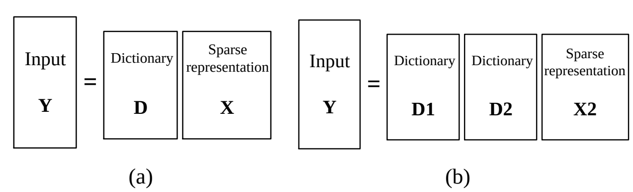

Figure 3. (a) Dictionary learning; (b) Deep dictionary learning

Considering the characteristics of the mechanical signal, we propose another deep stack dictionary learning (DSDL) algorithm, which is different from the method proposed by

Tariyal at al. (2016). In DSDL, the product of dictionary and representation from previous layer acts as the input of subsequent layer. The algorithm of DSDL diagram is displayed in Figure 4, where improved K-SVD (Rubinstein at al., 2008) as dictionary learning approach is integrated our method. Firstly, we apply improved K-SVD for original vibration signal and obtain sub-dictionaries and representation. In this process, in order to enhance the proximity between the measured signal y and its denoised vision z, it is better to add the log likelihood global term and the mathematical expression is as follows:

$$\arg\min_{z,\{x_i\}_i} \lambda \|y-z\|_2^2 + \sum_{i=1}^{N} \mu_i \|x_i\|_0 + \sum_{i=1}^{N} \|\mathbf{D}x_i - R_i z\|_2^2 \tag{8}$$

where first term is data fidelity term, the second and the third term are prior terms of the estimated signal. R_i is an operator that extracts patches from z, λ is the Lagrange multiplier. Although this optimization is nonconvex, we can utilize block coordinate minimization algorithm to solve the problem iteratively. Firstly, z is fixed, and the optimization problem is degraded to classical dictionary learning model, so we can obtain the x_i and $\mathbf{D}$ via the K-SVD algorithm.

$$\hat{x}_i = \arg\min_{x} \mu_i \|x\|_0 + \|\mathbf{D}x - R_i z\|_2^2 \tag{9}$$

Once all $\hat{x}_i$ are obtained, one can update z through solving the following optimization problem:

$$\arg\min_{z} \lambda \|z-y\|_2^2 + \sum_{i=1}^{N} \|\mathbf{D}\hat{x}_i - R_i z\|_2^2 \tag{10}$$

This is a simple quadratic problem, therefore, we can get a closed-form solution:

$$\hat{z} = (\lambda I + \sum_{i=1}^{N} R_i^T R_i)^{-1} (\lambda y + \sum_{i=1}^{N} R_i^T \mathbf{D}\hat{x}_i) \tag{11}$$

And updating $\hat{z}$ as the inputs of each successive later. The complete algorithm is summarized in algorithm 1.

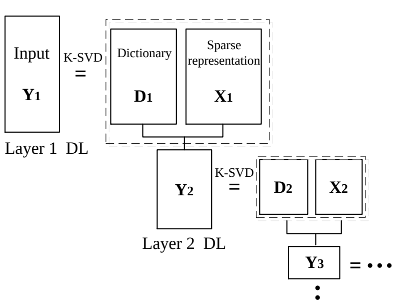

Figure 4. The structure of deep stack dictionary learning

Algorithm1: Deep Stack Dictionary Learning for extraction of periodic impulses

Input: original vibration signal y

Output: extracted fault feature $\hat{z}$

Initialization: number of iteration ***iter***, length of segments N, initial dictionary $\mathbf{D}$, the number of atoms M

For each layer dictionary learning

1. Extract Patches: Extract the patches from the examples and get data matrix.

2. Dictionary learning: Apply the improved K-SVD for input matrix, obtain sub-dictionary and corresponding coefficients.

3. Update inputs: sub-dictionary and representation are integrated as the input of successive layer.

Extract periodic impulses: $\hat{z}_{last\,layer}$

4. NUMERICAL SIMULATION

In this part, a numerical simulation is designed. When localized defects are generated on the rotating components, periodic impulses will be generated in the vibration signal. However, the periodic impulses are always submerged by heavy background noise. Therefore, the simulated signal is designed as follows:

$$y(t) = \sum_k A_k s(t - kT - \tau_k) + n(t) \tag{12}$$

Where y is the measured vibration signal, n is zero mean Gaussian noise, $s(t)$ denotes the impulses excited by a mechanical defect, where A_k is the amplitude of the kth impulse, T is the period of the impulsive signal, τ_k is the error from the position of the impulse and is a small random number. Generally, this term could be described by an exponentially decaying sinusoid with the following form:

$$s(t) = e^{-2\xi\pi f_r t} \sin(2\pi f_r \sqrt{1-\xi^2}\, t) \tag{13}$$

where f_r specifies the resonance frequency excited by the impact, ξ is the decay rate of the impulses.

Table 1. Simulation parameters

T	ξ	f_r	f	A_k
0.1	0.02	1500	10000	1.5

The parameters used for this model are described in Table 1. The waveforms of defect impulses and synthetic signal are illustrated in Figure 5, respectively. The signal to noise ratio (SNR) of simulation signal is -12.12, such that the fault impulses cannot be visualized from the final signal in Figure 5(b). The extracted impulsive signal via DSDL can be illustrated in Figure 6. The results demonstrate that our method has outstanding performance in extracting feature information from original vibration signal.

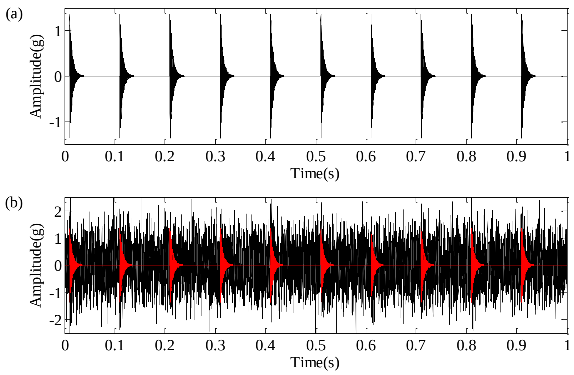

Figure 5. Simulated signal (a) defect impulses; (c) synthetic signal

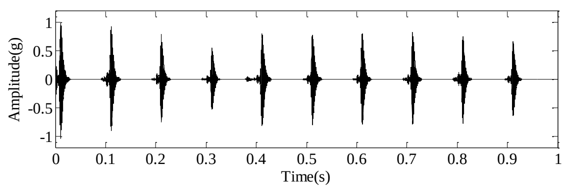

Figure 6. Feature signal by DSDL

5. EXPERIMENTAL VALIDATION

In this part, the vibration signal collected from locomotive bearing test rigs, which will be used to demonstrate the effectiveness of the DSDL.

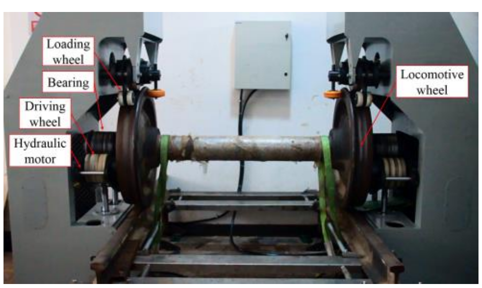

Figure 7. The locomotive bearing test bench

The experiments were made on a bearing test rig specialized for fault detection of locomotive bearings. The test rig consists of hydraulic motor, a driving wheel and a locomotive wheel and the overview is illustrated in Figure 7. To collect the vibration signal, a tri-axial PCB accelerometer with sensitivity of $100mVg^{-1}$ is mounted of the shaft end during the running process. An inner race fault was seeded and the ball-pass frequency of inner race (BPFI) is 80.675Hz. The sampling frequency is 76.8 kHz and the time length of the data is 0.5s.

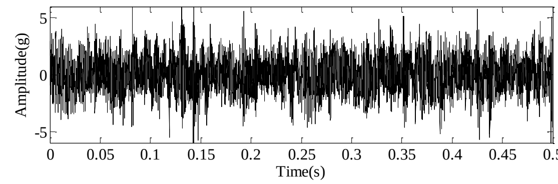

Figure 8. Raw signal of locomotive bearing

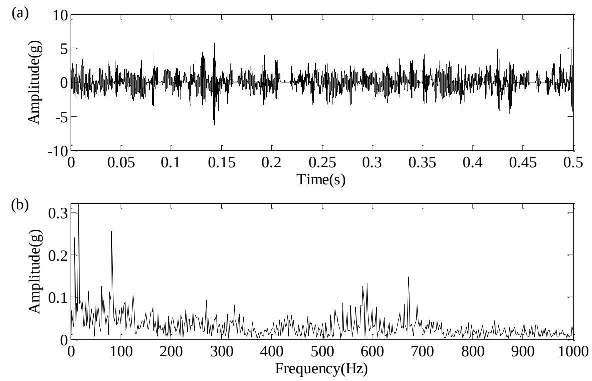

Figure 9. (a) Results of applying K-SVD to raw signal; (b) the envelope spectrum of (a)

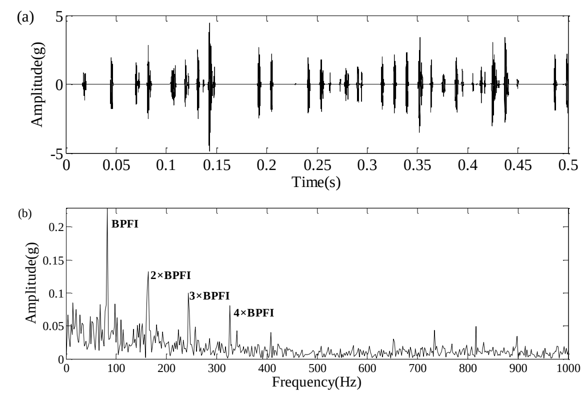

Figure 10. (a) Results of applying DSDL to raw signal; (b) the envelope spectrum of (a)

The raw signal is shown in Figure 8, in which we could not find any useful fault information due to heavy noise. Figure 9 and Figure 10 display the extracted periodic impulses and envelope spectrum through K-SVD and our algorithm, respectively. We found that the K-SVD cannot extract clear fault characteristic frequency. By comparison, our algorithm almost perfectly matches the original impulsive signals. The bearing inner race fault-related signatures and its harmonics are identified from Figure 10, which demonstrates that the proposed algorithm is effective and capable to extract the periodic defect impulses and remove the environment noise and other interference.

6. CONCLUSION

In the paper, a deep stack dictionary learning algorithm is proposed for fault diagnosis of rotating machinery. Utilizing

DSDL, we can obtain more expressive and robust feature which represent mechanical fault. Meanwhile, the effectiveness of our method is demonstrated by simulation signal and experimental data of locomotive bearing. Our future work will accelerate our algorithm and combine our algorithm with self-taught learning.

ACKNOWLEDGEMENT

This research is supported by the National Nature Science Foundation of China (Grant No. 51421004, 51405373), and the China Postdoctoral Science Foundation (2014M562400), which are highly appreciated by the authors.

REFERENCES

Lei, Y., Lin, J., He, Z., & Zi, Y. (2011). Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, vol. 25, pp. 1738-1749. doi:10.1016/j.ymssp.2010.12.011

Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, vol. 19.

Aharon, M., Elad, M., & Bruckstein, A. (2006). $ rm k $-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, vol. 54, pp. 4311-4322. doi:10.1109/TSP.2006.881199

Tropp, J. A., & Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, vol. 53, pp. 4655-4666. doi:10.1109/TIT.2007.909108

Rubinstein, R., Zibulevsky, M., & Elad, M. (2008). Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *Cs Technion*, vol. 40, pp. 1-15.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, vol. 313, pp. 504-507. doi: 10.1126/science.1127647

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. *InProceedings of the 25th international conference on Machine learning*. July 5-9. ACM. doi:10.1145/1390156.1390294

Tariyal, S., Majumdar, A., Singh, R., & Vatsa, M. (2016). Deep Dictionary Learning. *IEEE Access*, vol. 4, pp. 10096-10109. doi:10.1109/ACCESS.2016.2611583

Singhal, V., & Majumdar, A. (2017). Majorization Minimization Technique for Optimally Solving Deep Dictionary Learning. *Neural Processing Letters*, pp. 1-16. doi:10.1007/s11063-017-9603-9

BIOGRAPHIES

Jinyang Jiao is currently working toward the Ph.D. degree in mechanical engineering at the State Key Laboratory for Manufacturing System Engineering, Xi’an Jiaotong University, Xi’an, China. His research interest include machinery condition monitoring and intelligent fault diagnostics of rotating machinery.

Ming Zhao is currently a lecturer at School of Mechanical Engineering, Xi’an Jiaotong University, China. He received his BS, MS and PhD degrees from Xi’an Jiaotong University, in 2006, 2009, and 2013, respectively. He is working as a postdoctoral fellow and research associate in Center for Intelligent Maintenance Systems, University of Cincinnati, Ohio. His research interests include no-stationary signal processing, rotor dynamics and fault diagnosis of rotating machinery.

Jing Lin is a professor at State Key Laboratory for Manufacturing System Engineering, Xi’an Jiaotong University, China. He obtained his BSc, MSc and PhD degrees respectively in 1993, 1996 and 1999, all in mechanical engineering. He was working as a postdoctoral fellow and research associate from July 2001 to August 2003, respectively in University of Alberta, Canada, and University of Wisconsin-Milwaukee, USA. From September 2003 to December 2008, he was working as a research scientist at Institute of Acoustic, Chinese Academy of Science, under the Sponsorship of the Hundred Talents Program. He also obtained the National Science Fund for Distinguished Young Scholars in 2011. Now his research directions are mechanical system reliability, fault diagnosis and prognostics.