Bearing Fault Diagnosis Based on Deep Semi-supervised Small Sample Classifier

Hu Yongtao
School of Electrical Engineering and Automation
Henan Institute of Technology
Xinxiang, China
hythait@163.com

Gao Jinfeng
School of Electrical Engineering
Zhengzhou University
Zhengzhou, China
jfgao@zzu.edu.cn

Zhou Qiang
Intelligent Research Institute of Technology Center
Henan Weihua Heavy Machinery Co., Ltd
Xinxiang, China
Zhouqiang1000@163.com

Fan Zheng
School of Electrical Engineering and Automation
Henan Institute of Technology
Xinxiang, China
fanzheng1973@163.com

***Abstract*—In order to solve the difficult of fewer available samples in bearing fault diagnosis, a novel deep semi-supervised small-sample classifier (DSSC) based on restricted Boltzmann machine (RBM) is proposed and applied in the bearing fault diagnosis. DSSC can use little training samples with labels or a lot of unlabeled samples with a little of labeled samples to achieve an accurate classification. But it is a simple deep architecture and the number of input features should be small. So, fault feature extraction is conducted based on variational mode decomposition (VMD) and sample entropy. And the bearing fault diagnosis method is designed based on DSSC and VMD finally. Experiments are conducted to validate the effectiveness of the proposed method. The influence of size of training set and labeled samples are discussed. The experiment results show that the proposed method can make full use of massive unlabeled samples and little labeled samples and achieve fast and accurate bearing fault diagnosis, which can be extended to the online mechanical fault diagnosis.**

Keywords-bearing fault diagnosis; deep semi-supervised small-sample classifier; variational mode decomposition; sample entropy

I. Introduction

Bearings are essential components of mechanical equipment [1]. Because of the long-term operation under the harsh conditions of variable speed, heavy and alternating load, the bearings are inevitably damaged. If the equipment keeps running with a damaged bearing, the mechanical system will be broken down eventually, which results in enormous economic losses, or even serious casualties. Therefore, an effective method of bearing fault diagnosis is extremely valuable.

Feature extraction and fault classification are the two key steps of bearing fault diagnosis [2]. With the development of deep learning and its application in fault diagnosis, many experts have proposed intelligent fault diagnosis methods based on deep learning [3-5], which accept the original vibration signals to achieve the fault diagnosis. Tran *et al.* [6], proposed a new fault diagnosis method of reciprocating compressors valves based on deep belief network. Jing *et al.* [7] constructed a convolutional neural network (CNN) fault diagnosis model, which extracts features from frequency data of vibration signals and achieves the fault diagnosis of gears. Jiang *et al.* [8] proposed a deep learning fault diagnosis method based on stacked multilevel-denoising auto-encoders to improve the fault diagnosis accuracy. Wen *et al.* [9] proposed a deep transfer learning model based on sparse auto-encoder and achieve the intelligent fault diagnosis of bearings using the raw vibrational signals.

The method is intelligent because it combines feature extraction and fault classification into a fault diagnosis model and avoids human feature extraction. However, its training requires mounts of labeled samples to get the accurate model. This will be time consuming. And it is difficult to get the large number of labeled training samples in practice. What worse is that the structure and the parameters of the deep learning model are difficultly determined by experience. So, it is necessary to construct a semi-supervised small sample classifier, which can make full use of massive accessible unlabeled samples for unsupervised learning to exploit the fault features, and use little inaccessible labeled samples for supervised learning to achieve the accurate fault classification. In order to reduce the complexity of classifier structure and making the parameter selection easier, the traditional feature extraction based on time-frequency analysis is necessary. Variational mode decomposition (VMD) is a new self-adaptive signal processing method [10]. VMD can decompose the vibrational signal into a series of sub-signals, which is helpful to extract the accurate principal mode for fault feature extraction [11].

This work is supported by Key Scientific Research Projects Plan of Henan Higher Education Institutions (Grant No. 19B460001) and Science and Technology Research Project of Henan Province (Grant No. 182102210261)
978-1-7281-0860-5/19/$31.00 ©2019 IEEE

In this work, a semi-supervised small sample classifier based on DBN is proposed and used for bearing fault diagnosis. In order to reduce the complexity of classifier structure for fault diagnosis, fault feature is extracted by VMD and sample entropies. Experiments are carried out to prove the effectiveness of the proposed method.

The paper is organized as follows. Section II proposes the RBM based semi-supervised small sample classifier and introduces the feature extraction based on VMD and sample entropy. The bearing fault diagnosis model is constructed in Section III. In Section IV, bearing fault diagnosis experiments are performed. In Section V, conclusions are made.

II. RBM BASED SEMI-SUPERVISED SMALL SAMPLE CLASSIFIER

A. *Restricted Boltzmann Machine*

RBM is a stochastic neural network based on statistical mechanics [12], which consists of a visible layer $\mathbf{v}$ and a hidden layer $\mathbf{h}$. They are connected via symmetrically undirected weight $\mathbf{w}$. The structure of RBM is shown in Fig. 1.

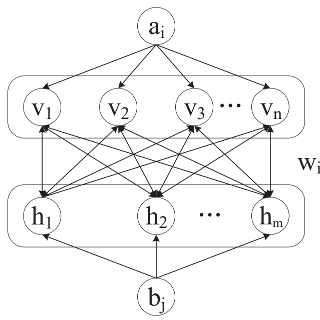

Figure 1. Structure of RBM

There is no connection between the visible layers and no connection between hidden layers. The main parameters of RBM are bias of visible layer $\mathbf{a}$, bias of hidden layer $\mathbf{b}$ and the weight between visible layer and hidden layer $\mathbf{w}$, which denote as $\boldsymbol{\theta} = \{\mathbf{a}, \mathbf{b}, \mathbf{w}\}$. The energy function of RBM can be defined as E:

$$E(\mathbf{v}, \mathbf{h} \mid \boldsymbol{\theta}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} v_i h_j - \sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j \quad (1)$$

Where v_i and h_j are the binary status of the *ith* visible neural cell and the *jth* hidden neural cell, a_i and b_j are the biases separately, w_{ij} is the connection weight between v_i and h_j, m and n are the number of neural cell in hidden and visible layer. The joint distribution of visible and hidden layers is defined as follow:

$$P(v,h) = \frac{1}{Z} e^{-E(v,h)} \quad (2)$$

where $Z = \sum_{v,h} e^{-E(v,h)}$ is a normalization constant, which ensures a normalized joint distribution. And the conditional probability of visible layers and hidden layers can be obtained by follows:

$$p(v_i = 1 \mid \mathbf{h}) = \delta(b_i + \sum_{j=1}^{m} w_{ij} h_j) \quad (3)$$

$$p(h_j = 1 \mid \mathbf{h}) = \delta(b_i + \sum_{i=1}^{n} w_{ij} v_i) \quad (4)$$

where $\delta(x) = \frac{1}{1+\mathrm{e}^{-x}}$ is the activation function. The parameters $\boldsymbol{\theta}$ can be simultaneously trained by minimizing the loss function. The loss function usually uses the mean square error of input and output of RBM.

B. *Deep Semi-supervised Small-sample Classifier (DSSC)*

A single RBM cannot effectively extract features from the input data and classify them. It is better to stacked several RBMs followed by a softmax classifier to construct a classifier with deep architecture. For fault diagnosis, it is difficult to get many labeled samples for training in practice. In fact, there are often a lot of unlabeled samples and a little of labeled samples. So, the classifier should accurately classify the data using a mount of unlabeled samples and just a little labeled sample. In order to solve this problem, a deep semi-supervised small sample classifier (DSSC) is designed as shown in Fig. 2. It consists of an input layer, three hidden layers, an output layer and a label layer. The input layer and the hidden layers constitute three RBMs.

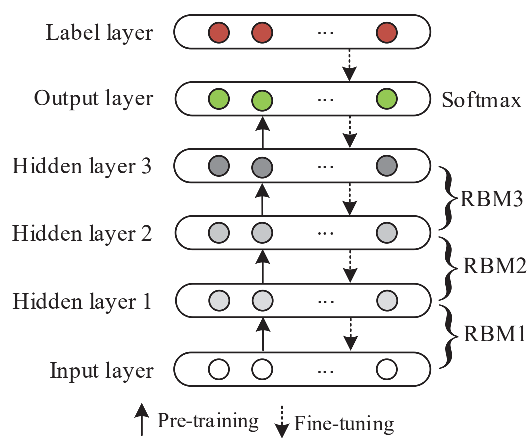

Figure 2. Deep semi-supervised small sample classifier

The input layer receives the input data and transmits them to the first hidden layer after encoding. The unit number of the input layer is set the same with the dimension of input data. The hidden layer 1 is the visible layer of RBM2. It receives the output of the first hidden layer, re-encoding and transmits it to the hidden layer 2. The hidden layer 2 is the visible layer of RBM3. It receives the output of the second hidden layer, re-encoding and transmits it to the third hidden layer. The unit number of hidden layers is big and difficult to determine when the unit number of input layer is very large. But has little effect on the classification performance when the unit number of input layer is small, and can be determined by experience. The output layer is a softmax classifier, which receive the features obtained by the third hidden layer and achieve the classification. The unit number of output layer is same with the classes number of all samples.

This layer by layer training procedure is an unsupervised learning, called pre-training. It performed as follow: the conditional probability of visible unit of RBM is calculated using (3) and the Gibbs sampling is employed for determining the visible unit state. Similarly, the hidden unit state of RBM can be obtained by (4). It can be considered as the reconstruction of visible unit in the previous layer. The gradient descent algorithm is then used to update the parameters $\boldsymbol{\theta}$. However, the classification accuracy is usually bad after pre-turning.

A label layer is stacked following the output layer to improve the classification accuracy. It takes a little of labeled samples to minimize the error between the output and the label by adjusting the parameters. This procedure is called fine-tuning, and the BP algorithms is utilized.

The model can be used for classification after pre-training and fine-tuning. The proposed deep semi-supervised small sample classifier has the following characteristics: 1) It has a simply deep structure and the number of the unit in each layer is easily to determined. 2) Its training includes unsupervised pre-training and supervised fine-tuning. So, it can make full use of mounts of unlabeled samples and a little of labeled samples to achieve accurate classification. 3) The unit number of input layer should small. So, feature extraction is necessary when used for fault diagnosis.

III. VMD and Sample Entropy Based Feature Extraction

A. VMD

VMD is a self-adaptive signal decomposition method. It can decompose a vibration signal into a series of modes. The bandwidth of all modes is limited and compact around a center frequency. It is a constrained variational problem in fact.

$$\begin{aligned}&\min_{u_k,\omega_k}\left\{\sum_k \left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\}\\&s.t.\quad \sum_k u_k = f\end{aligned} \tag{5}$$

where u_k is the mode and ω_k is the center frequency. The quadratic penalty term and Lagrange multiplier are employed to rendered the problem unconstrained. Then, the (5) can be rewritten as follow.

$$L(u_k,\omega_k,\lambda)=\alpha\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-j\omega_k t}\right\|_2^2+\left\|f-\sum_k u_k\right\|_2^2+\left\langle\lambda, f-\sum_k u_k\right\rangle \tag{6}$$

The u_k and ω_k can be obtained by the ADMM algorithm during each shifting process [13]. All modes in frequency domain can be gained from solutions as follow.

$$\hat{u}_k(\omega)=\frac{\hat{f}(\omega)-\sum_{i\neq k}\hat{u}_k(\omega)+\hat{\lambda}(\omega)/2}{1+2\alpha(\omega-\omega_k)^2} \tag{7}$$

The ω_k is updated as follows.

$$\omega_k=\frac{\int_0^\infty \omega|\hat{u}_k(\omega)|^2\,d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2\,d\omega} \tag{8}$$

The modes in time domain can be obtained by performing the inverse Fourier transform on $\hat{u}_k(\omega)$. The algorithm of VMD is detailed in [14].

B. Sample Entropy (SampEn)

SampEn is a new method of signal complexity measurement. It is an improvement of approximate entropy, has better consistency and does not depend on data length. The SampEn of a signal x is difined as follows.

$$SampEn(m,r,N)=-\ln\left[\frac{U^{m+1}(r)}{U^m(r)}\right] \tag{9}$$

$$U^m(r)=\frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}\sum_{j=1}^{N-m\tau}\frac{1}{N-(m+1)\tau}\Gamma(d_{ij}^m-r) \tag{10}$$

where m is the pattern length, r is the predetermined tolerance, N is the points number, τ is the time delay, $\Gamma(\cdot)$ is the Heaviside function.

$$\Gamma(x)=\begin{cases}1, & if\ x\le 0\\ 0, & if\ x>0\end{cases} \tag{11}$$

And d_{ij}^m is the distance between x_i^m and x_j^m. The detailed algorithm of SampEn can be found in [15].

The components and complexity of vibration signal are different when different faults occur in bearings, so the fault feature can be extracted easily by combining VMD with sample entropy. The steps of feature extraction based on VMD and sample entropy are as follows. Firstly, determine the VMD decomposition level k by experience and decompose the vibration signal into k sub-signals. Then, the SampEn of each sub-signal is calculated and k SampEn features are obtained for each vibration signal. Finally, the features are normalized and can be input to the proposed deep semi-supervised small sample classifier for fault diagnosis.

IV. FAULT DIAGNOSIS MODEL BASED ON VMD AND DSSC

In this section, the model of bearing fault diagnosis is designed based on VMD and DSSC as shown in Fig. 3. The steps of the bearing fault diagnosis are detailed as follow.

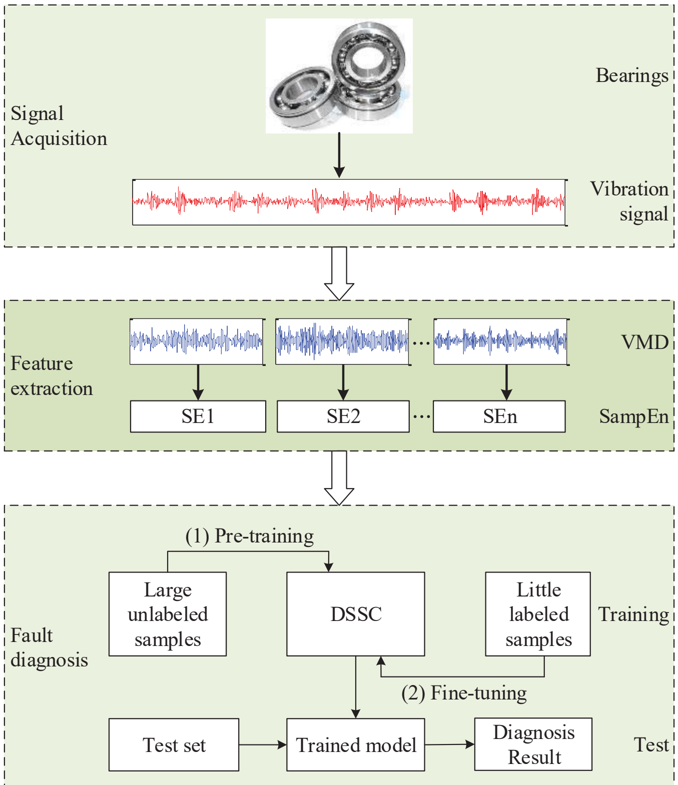

Figure 3. Bearing fault diagnosis model

Step 1: Collecting the vibration signals of bearings, including a lot of unlabeled samples and a little of labeled samples.

Step 2: Decomposing all samples into serials of modes. Then, the sample entropies of all modes are calculated and taken as the fault features matrix after normalized.

Step 3: Input the feature matrix of unlabeled samples pre-train the model.

Step 4: Input the feature matrix and labels of labeled samples to fine-tune the model. Then, the trained model is obtained.

Step 5: Extracting the fault features of test samples as step 2 and input to the trained model. The fault types of test samples can be recognized.

V. BEARING FAULT DIAGNOSIS EXPERIMENT

Experiments of bearing fault diagnosis are implemented to prove the validity of the proposed method. Bearing vibration signals are collected from the bearing fault simulation test-bed [16]. Single point faults are introduced in the ball, inner raceway and outer raceway by electro spark. The diameters include 0.007, 0.014 and 0.021 inches. So, there are 10 status, namely normal, ir007, ir014, ir021, 1007, b014, b021, or007, or014 and or021. They are labeled as 1-10 accordingly. Take 100 samples from each bearing for diagnosis experiment. Each sample has 1200 points. All samples are decomposed by VMD. The decomposition level is set to 6 according to the signal spectrum. And six sub-signals are obtained from each signal. Then, the sample entropies are calculated and normalized. The feature matrix with size of 1000×6 is obtained. Table1 is the feature of one sample of each class. The structure of DBN classifier is easily determined as $6\times6\times12\times12\times10$.

TABLE1 FEATURES OF ONE SAMPLE OF EACH CLASS

Sample class	**Feature**					
	1	***2***	***3***	***4***	***5***	***6***
1	0.0002	0.0024	0.1121	0.0027	0.0425	0.7635
2	0.2281	0.2895	0.8976	0.2012	0.0163	0.3384
3	0.1619	0.0832	0.0474	0.0226	0.0153	0.3530
4	0.2571	0.8353	0.1337	0.4081	0.0991	0.1740
5	0.0829	0.0437	0.0250	0.0120	0.0081	0.3403
6	0.0530	0.0335	0.0321	0.0399	0.2973	0.3361
7	0.0423	0.0142	0.0838	0.0311	0.0048	0.3362
8	0.7074	0.5737	0.2090	0.0848	0.0045	0.3371
9	0.0611	0.0044	0.0058	0.0233	0.0042	0.3366
10	0.8876	0.5679	0.4871	0.9305	0.0728	0.3363

A. Different Size of Training Dataset for Fault Fiagnosis

In order to verify the advantages of DSSC in small-sample recognition, the samples are divided into training sets and test sets. Take 5,10,15,20,25,30,35,40,45 and 50 samples from each status respectively as the training set, and 50 samples from each status as test set. So, there are 10 types experiments. Each experiment is repeated 20 times. The average and standard deviation of classification accuracy are used for verifying the performance of the method.

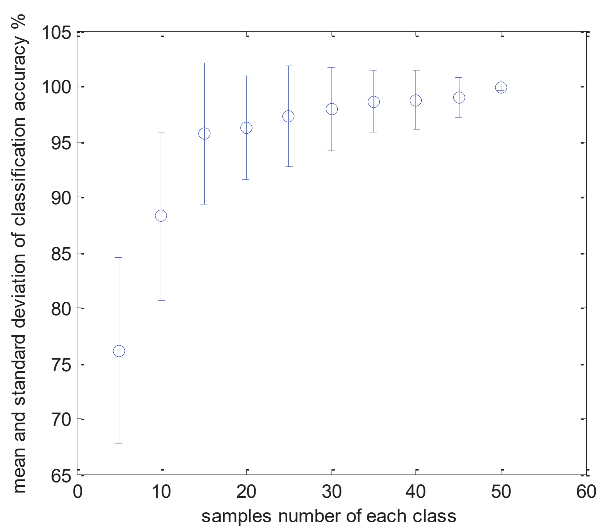

Figure 4. Classification accuracy and standard deviation of different size training set

Fig.4 shows the mean and standard deviation of classification accuracy for different training sets. It can be seen that the classification accuracy increases with the increase of the samples, and the standard deviation decreases. When the samples number in each category is too small, the classification accuracy is low and the standard deviation is large, so the test samples cannot be classified well. When there are 25 samples in each class (50% of the test set), the classification accuracy reaches 96.79% and the standard deviation is 4.94%. This is because with the increase of samples, the DSSC can learn more features and improve the classification accuracy and reliability.

Fig. 5 is the training accuracy curve when training set has 25 samples of each status. It shows that in the 100 iterations, the classification accuracy is not stable, and the iteration number needs to be increased. The prediction classification accuracy is 97.78%, and the training time is 11.2825s.

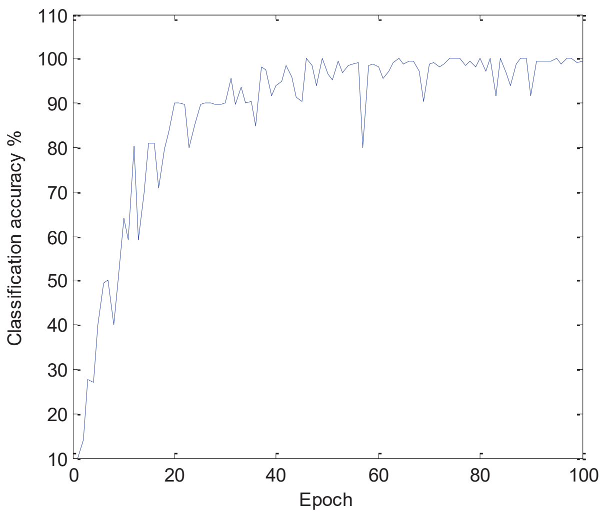

Figure 5. Training accuracy curves when each class has 25 samples

In order to improve classification accuracy, the iteration number of DSSC is set to 200, and the result is shown in Fig. 6. When the iteration number is 126, the classification accuracy reaches 100%, the prediction classification accuracy is 99.4%, and the training time is 18.9843s. It can be seen that increasing

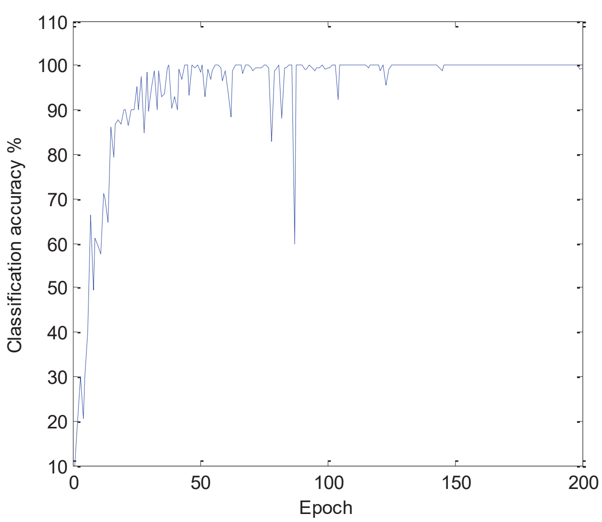

Figure 6. Training accuracy curves when each class has 25 samples and epoches is 200

the iteration number can improve the classification accuracy, but the time-consuming increases accordingly.

B. Different Size of Labeled Samples for Fault Diagnosis

To verify the effect of labeled samples on classification performance, 50 samples of each class are used for pre-training. Each class with 5,10,15,20,25,30,35,40,45 and 50 samples are used for fine-tuning. There are 10 types experiments totally and each experiment is repeated 20 times. Fig.7 is the mean and standard deviation of classification accuracy for different size of labeled samples. It shows that the classification accuracy increases rapidly with the increase of labeled samples. When the number of labeled samples is 20, that is, 40% of pre-training samples, the classification accuracy is 96.93%, and the standard deviation is 5.87%. Thereafter, with the increase of labeled samples, the classification accuracy increases gradually, while the standard deviation decreases gradually. Fig. 8 is the training accuracy curve for 20 labeled samples. When the number of iterations is 40, the classification accuracy reaches a stable 100%, the prediction classification accuracy is 100.00%, and the training time is 12.5518s.

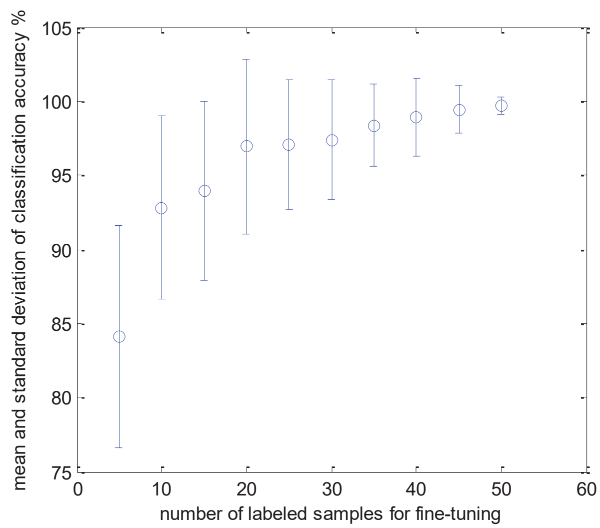

Figure 7. Classification accuracy and standard deviation of different labeled samples for fine-tuning

C. Discussion

Two experiments are were carried out in the study. In the first experiment, when the training set with labels reaches the 50% of the test set, it can be achieved accurate diagnosis. In the second experiment, when the labeled samples for fine-tuning reaches the 40% of training set, the classification accuracy reaches 100%. It is recommended that to improve the performance of the model, we can add appropriate samples in pre-training and using a small number of labeled samples for fine-tuning. Which is very impotent for the online fault diagnosis in practice.

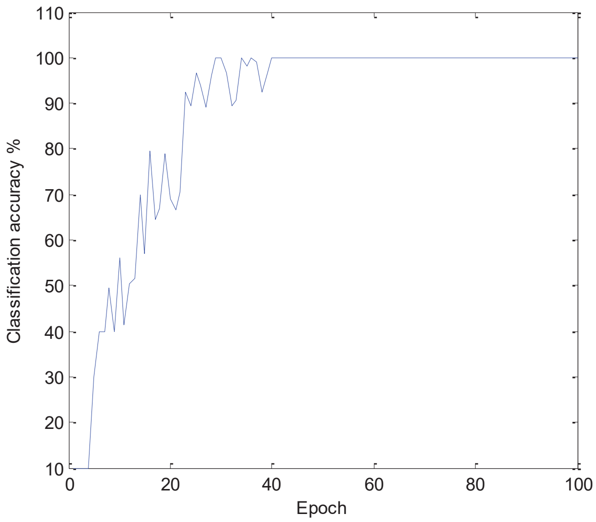

Figure 8. Training accuracy curves when 20 labeled samples used for fine-turning

VI. CONCLUSION

This paper proposes a new bearing fault diagnosis method based on deep learning. The major contributions of the paper are constructing a deep semi-supervised classifier for small sample classification, and applying it on the bearing fault diagnosis. The effectiveness of the method is verified by two experiments. The results show that the proposed classifier can make full use of a lot of unlabeled samples and a little of labeled samples to achieve an accurate classification. Which is very important for bearing fault diagnosis in the real world.

There are two points for future research. First, improved feature extraction methods can be used for feature extraction. Second, the method can be used for online mechanical fault diagnosis in practice.

REFERENCES

[1] L. W. Zhang, J. Lin, and R. Karim, “Sliding window-based fault detection from high-dimensional data streams,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 47, no. 2, pp. 289–303, Feb. 2017.

[2] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[4] P. Tamilselvan and P. F.Wang, “Failure diagnosis using deep belief learning based health state classification,” Rel. Eng. Syst. Safety, vol. 115, pp. 124–135, Jul. 2013.

[5] Y. G. Lei, F. Jia, J. Lin, S. B. Xing, and S. X. Ding, “An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data,” IEEE Trans. Ind. Electron., vol. 63, no. 5, pp. 3137–3147, May 2016.

[6] V. T. Tran, Thobiani F.; A. Ball, “An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks,” Expert. Syst. Appl., vol. 41, no. 9, pp. 4113–4122, 2014

[7] L. Y. Jing, M. Zhao, P. Li, X. Q. Xu, “A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox,” Measurement, vol. 111, no. 12, pp. 1-10, 2017

[8] G. Q. Jiang, H. B. He, P. Xie, Y. F. Tang, “Stacked Multilevel-Denoising Autoencoders: A New Representation Learning Approach for Wind Turbine Gearbox Fault Diagnosis,” IEEE Trans. Instrum. Meas., vol. 66, no. 9, pp. 2391-2402, 2017.

[9] L. Wen, L. Gao, X. Y. Li, “A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis,” IEEE Trans. Syst. Man Cyber.: Syst., vol. 49, no. 1, pp. 136-144, 2019

[10] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal Proces. vol. 62, pp. 531-544, 2014.

[11] M. Zhang, Z. Jiang, and K. Feng, “Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump,” Mech. Syst. Signal Proces., vol. 93, no. 15, pp. 460-493, 2017.

[12] G. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504-507, 2006.

[13] H. Liu, B. Song, H. Qin & “An Adaptive-ADMM algorithm with support and signal value detection for compressed sensing,” IEEE Sig. Proc. Lett., vol. 20, No. 4, pp. 315-318, 2013

[14] K. Dragomiretskiy, D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal. Proces., vol. 62, no. 3, pp. 531–544, 2014.

[15] J.S. Richman and J.R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” Am. J. Physiol. Heart. Circ. Physiol., vol. 278, pp. 2039–2049, 2000.

[16] Bearing Data Center. Seeded fault test data. http://csegroups.case.edu/bearingdatacenter/pages/download-data-file, 2013-10-15/2015-04-08