Bearing Fault Diagnosis based on Multi-task Learning

Wentao Mao*
School of Computer and Information Engineering
Henan Normal University
Xinxiang, CHINA, 453007
Email: maowt.mail@gmail.com

Jianliang He
School of Computer and Information Engineering
Henan Normal University
Xinxiang, CHINA, 453007
Email: hjlhdh@qq.com

Wushi Feng
School of Computer and Information Engineering
Henan Normal University
Xinxiang, CHINA, 453007
Email: fws910@qq.com

Siyu Tian
School of Computer and Information Engineering
Henan Normal University
Xinxiang, CHINA, 453007
Email: tianxiaosi77@qq.com

Abstract*—In recent years, machine learning techniques have been successfully applied to analyzing vibration signal in bearing fault diagnosis problems. However, the bottleneck to improve the diagnosis performance is the lack of valid domain knowledge about bearing fault in the classification or regression model, especially when the collected fault data is insufficient. Moreover, inadequate fault information will result in unstable diagnosis result. To solve this problem, a new bearing fault diagnosis method based on multi-task learning is proposed in this paper. The intuitive point is the bearing fault with similar crack size and loads can provide useful domain information for each other. By considering the diagnosis model on one bearing as a task, this method uses marginal spectrum by Hilbert-Huang Transform as features, and constructs multi-task learning model on multiple related diagnosis tasks with different fault types to improve their diagnosis performance simultaneously. Compared to the single diagnosis task, the proposed method shares the domain information between various fault types and then builds a more efficient diagnosis model. Experimental results on CWRU bearing data set show that, compared with some traditional machine learning-based diagnosis methods, the proposed method can effectively improve the diagnosis accuracy and robustness under different working conditions and it's especially suitable for dealing with lots of tasks with insufficient data. Even with insufficient features, the proposed method gets the accuracy from 77.50% to 94.17% , which indicates the multi-task diagnosis can get help from related bearing fault information.

Keywords: Bearing fault diagnosis, Multi-task learning, Hilbert-Huang Transform

I. INTRODUCTION

As the key components in rotating machinery, rolling bearings easily suffer from damages because of the tough working environments, which leads to machinery breakdown and economic loss. Therefore, it is important to develop reliable fault diagnosis method for bearings. In recent years, artificial intelligence techniques have been successfully applied in bearing fault diagnosis problems. How to raise the accuracy and stability of such methods has become a challenging problem in many real-world applications.

Because vibration signals can directly reflect working condition of rolling bearings, they are widely used in fault diagnosis methods with machine learning techniques. Generally speaking, there are two steps in these methods: (1) feature extraction process and (2) construction of diagnosis model. For (1), statistical characteristics of signals in time domain, frequency domain and time-frequency domain are widely used to describe bearings fault. For example, kurtosis and skewness can be used to describe the general trend of bearing degradation[1], while the incipient bearings fault is relatively easy to be represented by characteristic of frequency spectrum, especially in high-frequency parts [2]. For non-stationary signals, the energy spectrum obtained from Wavelet Packet Decomposition (WPD) and Empirical Mode Decomposition (EMD) are able to distinguish the difference between various fault types [3]. For (2), machine learning algorithms including Artificial Neural Network (ANN) and Support Vector Machine (SVM) have become a promising tool in bearing fault diagnosis. For example, Samanta et al.[4] using SVM and ANN to establish a diagnosis model based on bearing temporal fault feature. Caesarendra et al.[5] adopted relevance vector machine and logistic regression to estimate the bearings degradation level. Liu et al.[6] established an impact time-frequency dictionary for feature extraction, and used SVM for incipient fault diagnosis. According to the analysis above, we find two key factors for these methods: (1) the representative and discriminative ability of the feature; (2) the mapping relationship of machine learning method from feature to fault type. However, these methods still have some drawbacks: (1) most feature extraction methods work on statistical information whose feature representative ability are weak and could not select sensitive feature for specific learning task. (2) The bearings faults with same type may contain different data mode, which will be confused for direct use. (3) Single fault diagnosis task would raise the risk of over fitting.

We take the outer race fault as example. For outer race of a bearing, different types such as fracture, exfoliation and abrasion exist. Although these faults locate at same position, the modes of vibration signal may be quite different. Obviously, using these fault data to construct diagnosis model without discriminating these fault types will be hard to get more representative fault features. Furthermore, only learning the fault data of single fault type will restrict the diagnosis model to the specialized fault knowledge. If we can train several related and complementary diagnosis tasks simultaneously, the domain information about bearing fault will be enhanced and more representative features could be extracted.

Following the discussion above, we introduce multi-task learning(MTL) method to overcome the aforementioned drawbacks of bearing fault diagnosis. Multi-task learning [7] is a kind of machine learning method which can make full use of the specific domain information that is implied in the training

2166-5656/18/$31.00 ©2018 IEEE
DOI 10.1109/PHM-Chongqing.2018.00067

data of multiple related tasks, and improve the performance of these tasks simultaneously. In the past decade, multi-task learning method has been successfully applied in image processing [8], medical information processing [9] and other fields. Therefore, a new intelligent fault-diagnosis method is proposed in this paper based on multi-task learning. Firstly, marginal spectrum by Hilbert-Huang Transform [10-11] is used as basic features for multi-task learning. Second, one multi-task learning method, named $L_{2,1}$ Regularized Multi-task Joint Feature Selection method [12], is introduced as a basic modeling algorithm to train multiple diagnosis tasks with relatedness(e.g., the fault diagnosis for outer race and inner race under two different loads can be considered as two related tasks) by selecting joint discriminative features. As a result, the performance of each diagnostic task will be improved. To our best knowledge, there is no such work about bearing fault diagnosis based on multi-task learning.

II. A BRIEF INTRODUCTION TO HILBERT-HUANG TRANSFORM AND MULTI- TASK LEARNING

The multi-task fault diagnosis method introduced in this paper includes two main parts: (1) Hilbert-Huang Transform for marginal spectrum and (2) $L_{2,1}$ Regularized Multi-task Joint Feature Selection. Here we give a brief introduction for them.

A. Hilbert-Huang Transform

Most vibration signals are mixed with stationary or non-stationary signals plus background noises. Hilbert-Huang Transform (HHT) is the most promising time-frequency analysis technique for non-stationary vibration signal[13-14]. It consists of Empirical Mode Decomposition (EMD)[15] and Hilbert transform.

Given input vibration signal $x(t)$, EMD technique decomposes this signal into several intrinsic modal functions (IMFs):

$$x(t) = \sum_{i=1}^{k} c_i(t) + r_k(t) \tag{1}$$

where $c_i(t)$ is the i-th intrinsic modal function, and are $r_k(t)$ residuals. The Hilbert transform can be described as:

$$H[c(t)] = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{c(\tau)}{t-\tau} d\tau \tag{2}$$

Then we can get the analytic form of IMF defined as:

$$C_i^A(t) = c_i(t) + jc_i^H(t) = a_i(t) e^{j\theta_i(t)} \tag{3}$$

where $c_i^H(t)$ is calculate by Eq. (2), and $a_i(t) = \sqrt{c_i^2 + \left(c_i^H\right)^2}$,$\theta_i(t) = \arctan\left(c_j^H / c_i\right)$. Then the Hilbert marginal spectrum can be described as:

$$h_i(f) = \int h_i(f,t)\, dt = \int a_i^2(f_i, t) dt \tag{4}$$

where $h_i(f,t)$ represents the Hilbert spectral density, and $f_i(t) = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt}$. Then marginal spectrum of the whole signal can be described as:

$$H(f) = \sum_{i=1}^{n} h_i(f) \tag{5}$$

Using Eq. (5), we calculate every marginal spectrum of one sample of different fault signals and then set them as the input of multi-task learning.

B. Multi-Task Feature Selection

Multi-task learning is a kind of machine learning technique which is opposite to single-task learning. It can learn multiple tasks simultaneously. Therefore, multi-task learning can make full use of specific domain information implicitly in multiple related tasks, and improve the performance of every task. The differences between single-task and multi-task are shown in Figure 1.

This paper use a typical MTL method, named $L_{2,1}$

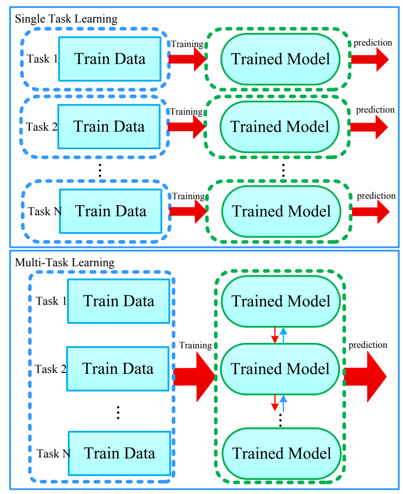


Fig. 1: The difference between single-task and multi-task

Regularized Multi-task Joint Feature Selection[12], to build the bearing fault diagnosis model. This method is good at capturing the task relatedness from multiple related tasks and then obtains the common features for all tasks. The loss function is shown as[12]:

$$\min_{W,c} \sum_{i=1}^{t} \left\|W_i^T X_i - Y_i\right\|_F^2 + \rho_1 \|W\|_{2,1} + \rho_2 \left\|W\right\|_F^2 \tag{6}$$

where X_i is the input matrix of the i-th task, Y_i is the corresponding label, W_i is the model for task i, the regularization parameter ρ_1 and ρ_2 controls group sparsity and L_2-norm penalty. Obviously, Eq. (6) uses $L_{2,1}$ loss as optimization target, and then gets sparse weight of features.

This paper employs the Nesterov's method [16-17] to solve the optimization problem of Eq. (6). Suppose that the optimization target can be seen as $\min_{\mathbf{x} \in G} g(x)$, and there are two sequence $\{\mathbf{x}_i\}$ and $\{\mathbf{s}_i\}$ for calculating this target where $\{\mathbf{x}_i\}$ is the sequence of solutions and $\{\mathbf{s}_i\}$ is the sequence of search points. The search point $\{\mathbf{s}_i\}$ is the combination of $\mathbf{x}_{i-1}$ and $\mathbf{x}_i$:

$$\mathbf{s}_i = \mathbf{x}_i + \alpha\left(\mathbf{x}_i - \mathbf{x}_{i-1}\right) \tag{7}$$

where α_i is the combination coefficient. The next solution $\mathbf{x}_{i+1}$ is calculated as:

$$\mathbf{x}_{i+1} = \pi_G \left(\mathbf{s}_i - \frac{1}{\gamma_i} g'(\mathbf{s}_i)\right) \tag{8}$$

where $\frac{1}{\gamma_i}$ is the step size, $\pi_G(\mathbf{v})$ is the Euclidean projection[18] of v onto the convex set G:

$$\pi_G(\mathbf{v}) = \min_{\mathbf{x}\in G} \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|^2 \tag{9}$$

The algorithm for solving equation (6) by the Nesterov's method is given in Algorithm 1.

Algorithm1. Nesterov's Method for Constrained Smooth Convex Optimization[12]

Input: $g(),\, g'(),\, G,\, L_0 > 0$
Output: $\mathbf{x}$
Initialize $\mathbf{x}_1 = \mathbf{x}_0, t_{-1} = 0, \gamma_0 = L_0$
for i=1 to do
Set $\alpha_i = \frac{t_{i-2}-1}{t_{i-1}}, \mathbf{s}_i = \mathbf{x}_i + \alpha(\mathbf{x}_i - \mathbf{x}_{i-1})$
 for j=0 to do
Set $\gamma = 2^j \gamma_{i-1}$
 Compute $\mathbf{x}_{i+1} = \pi_G\left(\mathbf{s}_i - \frac{1}{\gamma_i} g'(\mathbf{s}_i)\right)$
if $g(\mathbf{x}_{i+1}) \le g_{\gamma,s_i}(\mathbf{x}_{i+1})$ then
$\gamma_i = \gamma$, break
end if
end for
Set $t_i = \left(1 + \sqrt{1 + 4t_{i-1}^2}\right)/2$
if convergence then
$x = x_i$, terminate
 end if
end for

III. MTL-BASED BEARING FAULT DIAGNOSIS METHOD

Based on multi-task learning, this paper presents a new fault diagnosis method by optimizing common loss function from several related diagnosis tasks. The input of the propose method is the Hilbert-Huang marginal spectrum of raw vibration signal. Because the HHT intrinsic frequency is defined as the function of time, it is quite different from Fourier Transform that needs whole period of oscillation to define local frequency value. Therefore, the Hilbert-Huang marginal spectrum can reflect the local feature of raw signals more accurately, and then provide more clear information than frequency spectrum.

The total flowchart of the proposed method is shown in Fig. 2. Here the proposed method includes three parts:

(1) Collect the vibration signal of rotating bearings under different conditions. Then we get data set $\{\mathbf{x}_i = (x_1, x_2, \cdots, x_m), y_i\}_{i=1}^n$, where $\mathbf{x}^i$ is the i-th sample of vibration signal,y^i is the label of $\mathbf{x}^i$, and m is the point number of one sample, n is the sample number.

(2) Transform vibration signal to Hilbert-Huang marginal spectrum by using Eq. (5). In order to construct basic feature sets for multi-task learning, calculate marginal spectrum by Hilbert-Huang Transform. Then we get training set $\{\omega_i = (\omega_1, \omega_2, \cdots, \omega_k), y_i\}_{i=1}^n$, where ω^i is the i-th marginal spectrum of $\mathbf{x}^i$.

(3) Train several related tasks by using $L_{2,1}$ Regularized Multi-task Joint Feature Selection method. Here the related task is defined as the diagnosis task for same bearing under different work conditions or with different damage size. Input the training set of several tasks into Eq. (6) and calculate the best sparse weights of the multi-task model. Then employ the trained model to diagnose faults for the new bearing signals.

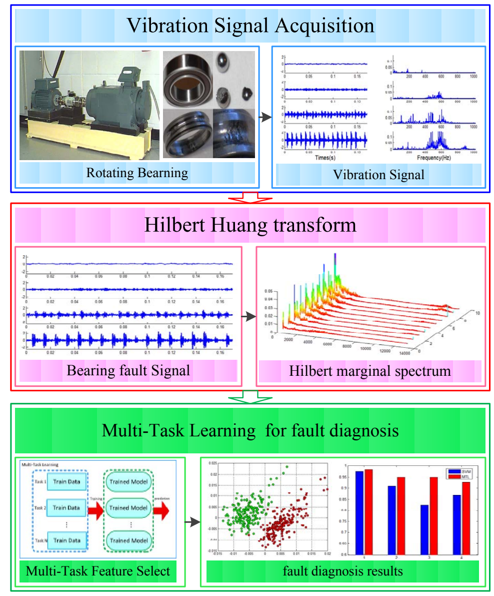


Fig. 2: The total flowchart of the proposed diagnosis method

IV. EXPERIMENTAL RESULTS

A. Data description and experiment setting

In this section, we introduce the bearing dataset provided by Case Western Reserve University [19] to run comparative experiments. The experiment equipment contains motors, acceleration sensors and power meters, etc. The faulted bearings are reinstalled into the motor system and the vibration signal is collected by accelerometer which is also reinstalled above the motor. The faults on bearings are seeded by electro-discharge machining (EDM). The vibration data was recorded for motor loads of 0 to 3 horsepower with the sample frequency of 12 kHz and 48 kHz. The fault ranging is from 0.007 inches to 0.021 inches in outer race, inner race and rolling element.

The main idea of this experiment is to verify the effectiveness of multi-task learning for fault diagnosis. For comparison, this paper uses SVM[20] training single diagnosis task. It's worth noting that the single task here means the fault diagnosis for one same bearing under different loads and with different crack size. We use LibSVM[20] as the implementation of SVM in this paper. And we extract 25-dimensional statistical features in temporal domain and frequency domain[1] as input, as shown in Table 1. To get the best performance in SVM, this paper choose the RBF kernel function and use cross-validation and grid search to determine the optimal regularization parameter C and kernel parameter g. The multi-tasking learning regularization parameters ρ_1 and ρ_2 in Eq. (6) are also chosen by means of cross-validation. The fan end (FE) data is used in all experiment. For vibration signal, the length of each sample is set 512, and the sample number per fault type is set 200. The experiment uses 70% of samples of each task as training set and the remaining 30% for test. The classification accuracy on the test set is calculated as the final result for comparison. The experimental environment is: CPU: i7-4790, Memory: 8G, Windows 7 operation system, MATLAB R2014a. It can verify the effectiveness of the multi-task learning method by comparing the two methods diagnostic accuracy.

TABLE I: 25-dimensional statistical feature

Method	Formula	Dimension
Maximum	Max(x)	1
Minimum	Min(x)	1
Absolute Average	$\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert$	1
Peak-to-Peak Value	Max(x)-Min(x)	1
Root Mean Square	$\left(\frac{1}{N}\sum\limits_{i=1}^{N}x_i^2\right)^{1/2}$	1
Average	$\frac{1}{n}\sum_{i=1}^{n}x_i$	1
Standard Deviation	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$	1
Skewness	$\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{3/2}}$	1
Kurtosis Factor	$\frac{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma}\right)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2\right)^2}$	1
Variance	$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$	1
Wave Factor	$\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert^2}}{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert}$	1
Peak Factor	$\frac{\max(x)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i)^2}}$	1
Change Coefficient	$\frac{\bar{x}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}}$	1
Skewness Coefficient	$\frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right)^3$	1
Kurtosis Coefficient	$\frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right)^4$	1
Clearance Factor	$\frac{Max(x)}{\frac{1}{n}\sum_{i=1}^{n}(x_i)^2}$	1
Impulse Factor	$\frac{\max(\lvert x_i\rvert)}{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$	1
Frequency Center	$\frac{\sum_{j=1}^{n}f_jX_j}{\sum_{j=1}^{n}X_j}$	1
Frequency RMS	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}$	1
Frequency Standard Deviation	$\sqrt{\frac{\sum_{j=1}^{n}(f_j)^2X_j}{\sum_{j=1}^{n}X_j}}$	1
Frequency Spectrum Partition Summation	$\sum_{j=\frac{K+m(k-1)}{K}}^{mk/K}f_j$	1

We set up 6 groups of experiment for multi-task learning. This setup is based on the following hypothesis: (1) The fault types at same position(i.e., inner race fault, outer race fault and so on) with different crack size have inner relatedness; (2) The fault types at same position under different loads have certain relatedness. As a result, three verification experiments are set for each hypothesis, and the comparative study between multi-task learning method and single task learning method is conducted. The task settings are shown in Table 2.

TABLE II: Experimental setup of multi-task learning

	Fault type	Working conditions	Task	Task number
Test 1	Ball fault and outer fault	Load 1 48kHz FE	0.007 inch 0.014 inch 0.021 inch	3
Test 2	Ball fault and outer fault	Load 2 48kHz FE	0.007 inch 0.014 inch 0.021 inch	3
Test 3	Ball fault and outer fault	Load 3 48kHz FE	0.007 inch 0.014 inch 0.021 inch	3
Test 4	Ball fault and outer fault	48kHz FE 0.021 inch	Load 0 Load 1 Load 2 Load 3	4
Test 5	Ball fault and inner fault	48kHz FE 0.021 inch	Load 0 Load 1 Load 2 Load 3	4
Test 6	Outer fault and inner fault	48kHz FE 0.021 inch	Load 0 Load 1 Load 2 Load 3	4

B. Comparative results

We firstly check the performance on Test 1-3. The diagnosis results on test set are shown in Table 3-5. For better illustration, Fig. 3 also provides the comparative performance between multi-task and single task (SVM) learning on Test 1-3.

From Fig.3, the accuracy (red) of multi-task learning is higher than the accuracy (blue) of single-task learning on three tests. It mean that the proposed method gets the higher classification accuracy than SVM from the related task, which indicates that multi-task learning can enhance the performance of all tasks simultaneously.

The diagnosis results on Test 4-6 in Table 2 are shown in Fig.4. Detailed comparative results for MTL and SVM are also shown in Table 6-8.

TABLE III: Classification accuracy of ball fault and outer race fault with three crack size under Load 1 (Test 1)

	SVM	MTL
0.007 inch	100%	100%
0.014 inch	97.50%	**99.17%**
0.021 inch	90.00%	**95.83%**

TABLE IV: Classification accuracy of ball fault and outer race fault with three crack size under Load 2 (Test 2)

	SVM	MTL
0.007 inch	98.33%	**99.17%**
0.014 inch	96.67%	100%
0.021 inch	77.50%	**95.00%**

TABLE V: Classification accuracy of ball fault and outer race fault with three crack size under Load 3 (Test 3)

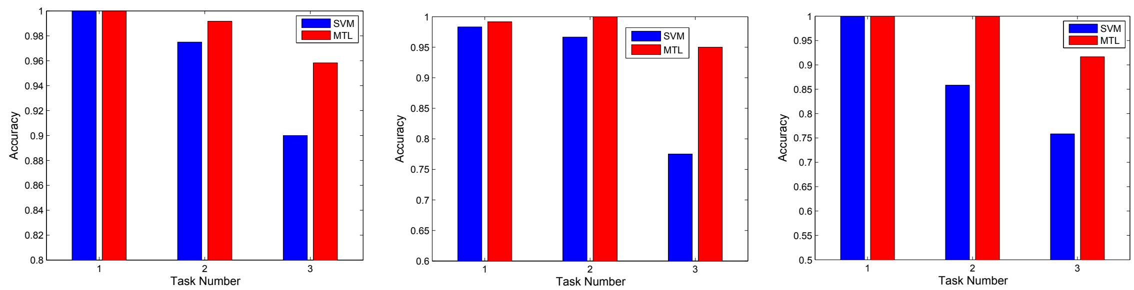
	SVM	MTL
0.007 inch	100%	100%
0.014 inch	85.83%	**100%**
0.021 inch	75.83%	**91.67%**

From the result of Fig.4, the red colors are also higher than blue colors in each sub-figure, which are also represent multi-task learning and single-task learning method. It also indicates that the proposed multi-task learning gets the highest classification accuracy than single-task learning from the related task and enhances the performance of all tasks simultaneously.

Besides four cases of 0.007inch and load 0(in Table 3, 5, 6, 7), MTL all get higher classification accuracy than SVM,

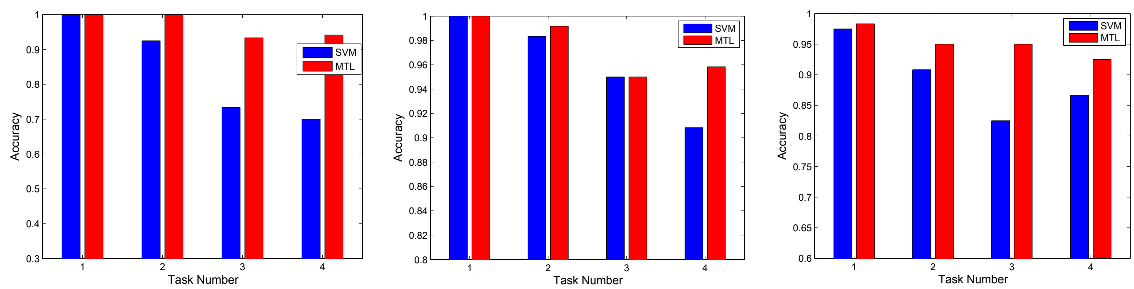
TABLE VI: Classification accuracy of ball fault and outer race fault under four loads (Test 4)

	SVM	MTL
Load 0	100%	100%
Load 1	92.50%	**100%**
Load 2	73.33%	**93.33%**
Load 3	70.00%	**94.17%**



(a) (b) (c)

Fig. 3: Comparative performance on (a) Test 1, (b) Test 2 and (c) Test 3



(a) (b) (c)

Fig. 4: Comparative performance on (a) Test 4, (b) Test 5 and (c) Test 6

TABLE VII: Classification accuracy of ball fault and inner race fault under four loads (Test 5)

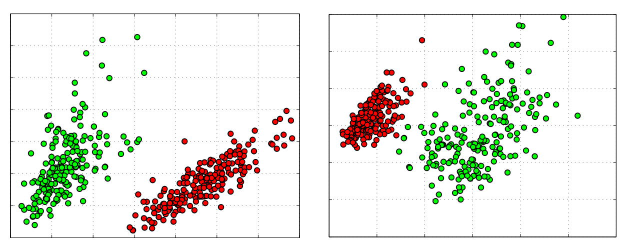
	SVM	MTL
Load 0	100%	100%
Load 1	98.33%	**99.17%**
Load 2	95.00%	95.00%
Load 3	90.83%	**95.83%**

TABLE VIII: Classification accuracy of outer race fault and inner race fault under four loads (Test 6)

	SVM	MTL
Load 0	97.50%	**98.33%**
Load 1	90.83%	**95.00%**
Load 2	82.50%	**95.00%**
Load 3	86.67%	**92.50%**

which indicates multi-task learning can improve the diagnosis performance by utilizing fault domain information from related diagnosis tasks, especially with small-scale sample size. We also notice that load 3 and 0.021inch tend to raise more apparent comparison between MTL and SVM. Under these two conditions, single task learning using SVM generally gets lowest diagnosis accuracy, however, after introducing multi-task learning which can get help from other fault types, the accuracy gets much increase. This phenomenon also demonstrates different fault types exist some relatedness, and this kind of relationship can help to enhance the diagnosis performance for each other.

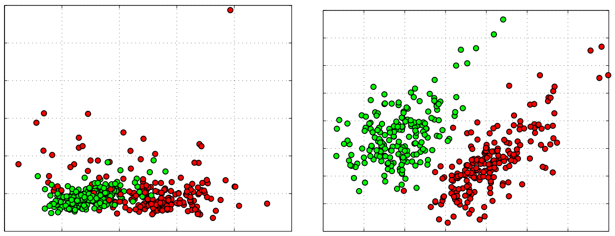
To verify the separability of the obtained features from the multi-task learning, we use Principal Components Analysis (PCA) to visualize these features in Fig. 5. As a comparison, the statistical features from Table 2 are also represented with different color.



(a) (b)

Fig. 5: Visualization of feature distribution by PCA for Task 1 in Test 1 with (a) traditional statistical features and (b) joint features by multi-task learning to diagnose ball fault(red point) and outer race fault(green point).

In Fig. 5, different color features locate enough far to each other, which brings good separability for diagnosis. As a result, using single task learning(SVM) and multi-task learning can both get good classification accuracy, which keeps line with the results in Table 3. The classification accuracy of both methods are both 100%. We also visualize these features for Task 3 in Test 1 by the same way, as shown in Figure 6.



(a) (b)

Fig. 6: Visualization of feature distribution by PCA for Task 3 in Test 1 with (a) traditional statistical features and (b) joint features by multi-task learning to diagnose ball fault(red point) and outer race fault(green point)

In Figure 6(a), we observe that the features of two classes are mixed together, which indicates that the representative and discriminative ability of the common statistical feature are not good enough. In this case, the corresponding classification accuracy is 90%. However, in Figure 6(b), two classes have good separability by multi-task learning, and this distribution brings 95.83% classification accuracy. Therefore, multi-task learning can learn automatically the joint features which can represent the fault domain information more accurately than the traditional features.

V. Conclusions

In this paper, a new bearing fault diagnosis method based on multi-task learning is proposed. Specifically speaking, this method includes obtaining feature representations by Hilbert-Huang transform and constructing bearing fault diagnosis model with $L_{2,1}$ Regularized Multi-task Joint Feature Selection. This method tries to improve the diagnostic performance by training multiple related tasks for different fault types simultaneously. From the experimental results, the conclusions are as follows:

(1) In terms of feature extraction, the Hilbert-Huang Transform marginal spectrums are more discriminative than statistical feature in time and frequency domain. They are more suitable to perform as the input data of multi-task learning.

(2) In terms of diagnosis model, the used $L_{2,1}$ Regularized Multi-task Joint Feature Selection method can utilize specific domain information implied in the training data of different fault types effectively. As a result, multi-task learning can get help from other related (not same) tasks to get a better diagnosis result.

In the next work, our research interest will focus on deep learning-based diagnosis model. Considering the deep features can be extracted automatically with more representative ability, some new deep multi-task learning methods will be developed and applied to bearing fault diagnosis problems.

Acknowledgements

This work was supported by the National Natural Science Foundation of China(No. U1704158, U1604154), China Postdoctoral Science Foundation Special Support(No. 2016T90944), the funding scheme of University Science & Technology Innovation in Henan Province(15HASTIT022) and the foundation of Henan Normal University for Excellent Young Teachers(No.14YQ007).

References

[1] T. W. Rauber, F. D. A. Boldt, and F. M. Varejão. "Heterogeneous Feature Models and Feature Selection Applied to Bearing Fault Diagnosis." IEEE Transactions on Industrial Electronics, vol. 62, no. 1, pp. 637-646, 2015.

[2] A. SoualhiK. Medjaherand N. Zerhouni. "Bearing Health Monitoring Based on HilbertCHuang Transform, Support Vector Machine, and Regression." IEEE Transactions on Instrumentation & Measurement, vol. 64, no. 1, pp. 52-62, 2014.

[3] W. Mao, L. He, J. Wang, et al. "Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine." Mechanical Systems & Signal Processing, vo. 83, pp. 450-473, 2017.

[4] B. Samanta, and C. Nataraj. "Use of particle swarm optimization for machinery fault detection," Engineering Applications of Artificial Intelligence, vol. 22, no. 2, pp.308-316, 2009.

[5] W. Caesarendra, A. Widodo, and B. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," Mechanical Systems & Signal Processing, vol. 24, no. 4, pp.1161-1171, 2010.

[6] R. Liu, B. Yang, X. Zhang, et al. "Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. Mechanical Systems and Signal Processing, 2016, 75(15): 345-370.

[7] Scholköpf, B, J. Platt, and T. Hofmann, "Multi-Task Feature Learning." in Proceedings of Conference on Advances in Neural Information Processing Systems, MIT Press, 2006, pp. 41-48.

[8] M. Julien, F. Bach, and J. Ponce, "Sparse Modeling for Image and Vision Processing." Foundations & Trends in Computer Graphics & Vision, vol. 8, no. 2, pp. 85-283, 2014.

[9] F. Liu, C. Y. Wee, et al. "Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification." Neuroimage, vol. 84, pp. 466-475, 2014.

[10] N. E. Huang, Z. Shen, S. Long, et al., "The Empirical Mode Decomposition and Hilbert Spectrum for nonlinear and non-stationary time series analysis, " in Proc. R. Soc. London A, 1998, pp.903C995.

[11] N. E. Huang, Z. Shen, S. Long, "A new view of nonlinear water waves: The Hilbert Spectrum," Annual Review of Fluid Mechanics, vol.61, pp.417C457, 1999.

[12] J. Liu, S. Ji, and J. Ye. "Multi-Task Feature Learning Via Efficient $L_{2,1}$-Norm Minimization. " in Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2009, pp. 339-348.

[13] H. Li, Y. Wang, and Y. Ma, "Ensemble empirical mode decomposition and Hilbert-Huang transform applied to bearing fault diagnosis," in 3rd International Congress on Image and Signal Processing (CISP), vol.7, 2010, pp.3413-3417.

[14] R. Yan, and R. X. Gao, "HilbertCHuang transform-based vibration signal analysis for machine health monitoring," IEEE Trans. on Instrum. Meas., vol. 55, no. 6, pp.2320C2329, 2006.

[15] N. E. Huang, Z. Shen, S. Long, et al., "The Empirical Mode Decomposition and HilbertSpectrum for nonlinear and non-stationary time series analysis, " in Proc.R. Soc. London A, 1998, pp.903C995.

[16] A. Nemirovski. "Efficient methods in convex programming." Lecture Notes, 1994.

[17] Y. Nesterov. "Introductory Lectures on Convex Optimization: A Basic Course." Kluwer Academic Publishers, 2003.

[18] S. Boyd, and L. Vandenberghe. "Convex Optimization." Cambridge University Press, 2004. Chapter 8.1, page 397.

[19] X. S. Lou, and K. A. Loparo. "Bearing fault diagnosis based on wavelet transform and fuzzy inference." Mechanical Systems & Signal Processing, vol. 18, pp. 1077C1095, 2004.

[20] C. C. Chang, and C. J. Lin. "LIBSVM: A library for support vector machines." http://www.csie.ntu.edu.tw/cjlin/libsvm/index.html, 2001. (Acccessed February 2016).