

# Application of Xgboost Feature Extraction in Fault Diagnosis of Rolling Bearing

Xingang WANG\*, Chao WANG

Research Center of Mechanical Kinetics and Reliability, Northeastern University, Qinhuangdao, China

\*Corresponding Author: Xingang WANG, Research Center of Mechanical Kinetics and Reliability, Northeastern University, Qinhuangdao, 066004, China; xgwang@neuq.neu.cn

## Abstract:

Due to the difficulty that excessive feature dimension in fault diagnosis of rolling bearing will lead to the decrease of classification accuracy, a fault diagnosis method based on Xgboost algorithm feature extraction is proposed. When the Xgboost algorithm classifies features, it generates an order of importance of the input features. The time domain features were extracted from the vibration signal of the rolling bearing, the time-frequency features were formed by the singular value of the modal components that were decomposed by the variational mode decomposition. Firstly, the extracted time domain and time-frequency domain features were input into the support vector machine respectively to observe the fault diagnosis accuracy. Then, Xgboost algorithm was used to rank the importance of features and got the accuracy of fault diagnosis. Finally, important features were extracted and the extracted features were input into the support vector machine to observe the fault diagnosis accuracy. The result shows that the fault diagnosis accuracy of rolling bearing is improved after important feature extraction in time domain and time-frequency domain by Xgboost.

**Keywords:** fault diagnosis;rolling bearing;xgboost;feature extraction;support vector machine

## 1. Introduction

Nowdays, the requirement of safety and reliability of mechanical products is higher and higher<sup>[1]</sup>. The fault diagnosis research on the most widely used rolling bearing can improve the safety of mechanical operation.

Fault diagnosis of rolling bearing usually consists of feature extraction and fault identification. The commonly used feature extraction methods include time domain, frequency domain and time-frequency domain. Zhang et al. <sup>[2]</sup> diagnosed the fault of rolling bearing by extracting the time domain features of the vibration acceleration signal of the rolling bearing, such as kurtosis. Kurtosis is sensitive to fault classification <sup>[3]</sup>, while some other time domain features are not sensitive to fault classification, which may lead to low accuracy of fault recognition due to too many feature dimensions. Empirical Mode Decomposition (EMD) can adaptively decompose the signal into a series of intrinsic mode functions from high frequency to low frequency. However, this method is easy to mode mixing. In 2014, Dragomiretskiy et al. <sup>[4]</sup> proposed a new adaptive signal processing method-Variational Modal Decomposition (VMD). The VMD

method iteratively searched the optimal solution of the Variational model to determine the frequency center and bandwidth of each intrinsic modal component, which can realize the frequency domain subdivision of signals and the effective separation of each signal component. However, the impact of penalty parameter  $\alpha$  and decomposition number K should be considered in VMD decomposition. Envelope entropy can be used to represent the sparse characteristics of signals. After signal decomposition, if there are more noise in signal components, the envelope entropy will be smaller. The optimal combination (K,  $\alpha$ ) can be obtained by taking the minimum value of envelope entropy as the objective function <sup>[5]</sup>. Wang et al. <sup>[6]</sup> used particle swarm optimization algorithm to optimize the VMD method to obtain the optimal (K,  $\alpha$ ) combination to decompose the signal. Liu et al. <sup>[7]</sup> further improved it by using particle swarm optimization based on genetic variation. Extracting the features that can easily distinguish different fault types are the most important for fault diagnosis <sup>[8]</sup>. Swarm intelligence optimization algorithms tend to fall into the local optimal solution, and the modal components obtained by VMD method decomposition do not completely represent fault characteristics (fault

frequency is not obvious for other noise information and interference).

At present, fault recognition systems based on machine learning algorithms are widely used. Support Vector Machine (SVM) method [11] has been widely used in many fields such as fault diagnosis and pattern recognition [12]. In terms of nonlinear problems, SVM introduces penalty parameters and kernel functions to transform them into linear problems in high dimensional space, so as to achieve effective classification. Boosting decision tree is an extensive and efficient integrated learning method. As a new method of boosting decision tree, Xgboost algorithm has been widely applied in economy and medical [13], and has achieved good results.

## 2. Xgboost Algorithm

The Xgboost algorithm is short for eXtreme Gradient Boosting. It is a c++ implementation of the gradient boosting machine algorithm developed by Dr. Chen Tianqi of Washington University [14]. The biggest advantage of Xgboost is that it can automatically take advantage of the CPU's multiple threads to run in parallel, while also improving the algorithm to improve the accuracy. Compared with Xgboost algorithm, the traditional Gradient Boosting Decision Tree (GBDT) algorithm only uses the information of the first derivative and the residual of the first n-1 tree to train the nth tree, which is difficult to realize distribution. Xgboost performs a second order Taylor expansion of the loss function (the objective function) and adds a regular term to the loss function to optimize the overall solution, so it can be used to weigh the decline of the objective function and the complexity of the model to avoid overfitting the model.

Suppose the model has k decision trees.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \Phi \quad (1)$$

The loss function (objective function) can be defined as

$$L = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

where  $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$ , T is the number of leaves, w is the weight of leaves. And

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3)$$

Then the loss function can be expressed as

$$L^{(t)} = \sum_{i=1}^n l(\hat{y}_i^{(t-1)}, y_i + f_t(x_i)) + \Omega(f_t) \quad (4)$$

Taylor expansion of the loss function can be written as

$$L^{(t)} \approx \sum_{i=1}^n [l(\hat{y}_i^{(t-1)}, y_i + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i))] + \Omega(f_t) \quad (5)$$

$$\text{where } g_i = \partial_{\hat{y}_i^{(t-1)}} l(\hat{y}_i^{(t-1)}, y_i) \quad (6)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(\hat{y}_i^{(t-1)}, y_i) \quad (7)$$

Remove the constant term, and the loss function as

$$\tilde{L}^{(t)} = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (8)$$

The algorithm is transformed from traversing the sample to traversing the leaf node. Let  $I_j = \{i | q(x_i) = j\}$  be defined as the j-th leaf point.

$$\begin{aligned} \tilde{L}^{(t)} &= \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\ &= \sum_{j=1}^T [ \sum_{i \in I_j} g_i w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2 ] + \gamma T \end{aligned} \quad (9)$$

let  $G_j = \sum_{i \in I_j} g_i, H_j = \sum_{i \in I_j} h_i$ , take the derivative of (9) and

set it to 0 to get the minimum of the objective function.

$$w_j = \frac{G_j}{H_j + \lambda} \quad (10)$$

Substitute the optimal solution into the objective function

$$\tilde{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (11)$$

Xgboost is a greedy algorithm. Each training will add segmentation to the existing leaf nodes. Suppose L and R are the set of left and right nodes after segmentation.

Information gain is as equation (12)

$$\text{Gain} = \frac{1}{2} \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (12)$$

As can be seen from the above formula, Xgboost uses a certain value after division to subtract a certain value before division. In order to control the growth of the tree, the algorithm adds a threshold value  $\gamma$ . Only when the information gain is greater than the threshold value, the algorithm allows the nodes to segment. It is also the coefficient of the leaf nodes in the regular term, so it is equivalent to pre-pruning the tree model while optimizing the loss function.

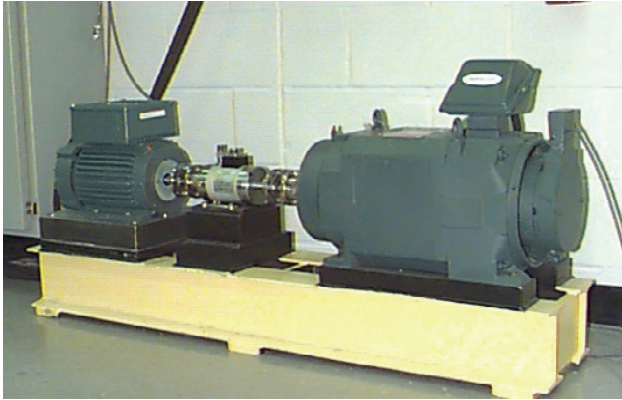
The Xgboost algorithm uses the decision tree model to classify the attributes of child nodes, and expresses the feature importance to the current classifier through the product of weight and leaf nodes. Because Xgboost uses the idea of integration (grouping multiple weak classifiers into one classifier), it adds up the importance of the same feature to multiple trees (sorted by different attributes) and gets the F score for the importance of the feature throughout the classification process.

## 3. Feature Extraction

### 3.1 Data

The vibration acceleration signal data [15] obtained by simulating different states of rolling bearings in the electrical engineering laboratory of Case Western Reserve University (CWRU) is taken as the object (the test rig picture is shown

in fig.1). It is known that SKF deep groove ball bearing model 6205-2RS JEM was selected for the experiment, fault points with fault diameters of 0.178mm were introduced by manual discharge operation. The sampling frequency of vibration signal is 12kHz, and the motor speeds are 1797r/min. In this paper, the early vibration signal data are used to research.



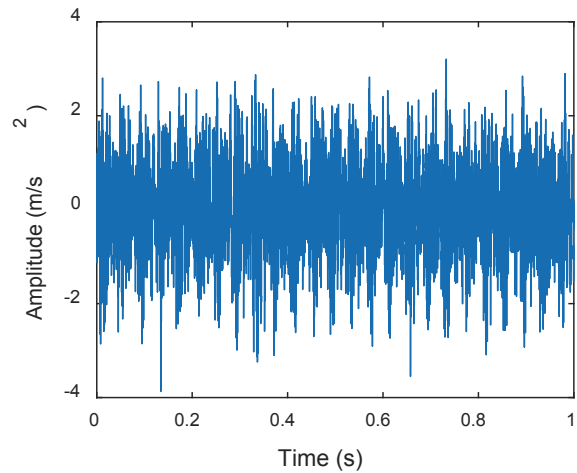
**Figure 1** The test rig of CWRU. As shown in Figure 1 above, the test stand consists of a 2 hp motor (left), a torque transducer/encoder (center), a dynamometer (right), and control electronics (not shown). The test bearings support the motor shaft. Single point faults were introduced to the test bearings using electro-discharge machining with fault diameters of 7 mils, 14 mils, 21 mils, 28 mils, and 40 mils (1 mil=0.001 inches).

### 3.2 Data Analysis

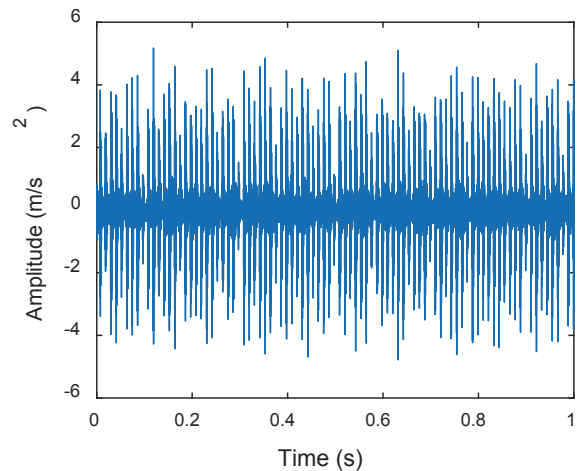
It is assumed that the four states of the rolling bearing are healthy, bearing element fault, outer ring fault and inner ring fault, and the four states are represented by 0, 1, 2 and 3 respectively. Fig.2 to 5 are respectively the vibration signals of the bearing in four states (the data was normalized). It can be seen that the fault type of the bearing can hardly be distinguished by the time domain signals of the rolling bearing. Fig.6 shows a series of modal components and corresponding spectrum obtained by VMD ( $K=6, \alpha=2000$ ) of no.105 data and bearing signals at the driver end of the test bench. The envelope spectrum of the first modal component was obtained as shown in fig.7. It was found that the inner ring fault frequency could be obtained, but it was not obvious. Fig.8 is the envelope spectrum obtained by enveloping the second modal component, and it can be seen that the fault characteristic frequency was obtained within the allowable error range. After enveloping all the modal components, it was found that only the second modal component had obvious fault characteristics, so it is necessary to extract the features. The failure frequency calculation formula of the inner ring <sup>[16]</sup> is as follows

$$f_{ir} = \frac{N}{2} \left(1 + \frac{d}{D} \cos \theta\right) f_r; \text{ Inner ring failure frequency (13)}$$

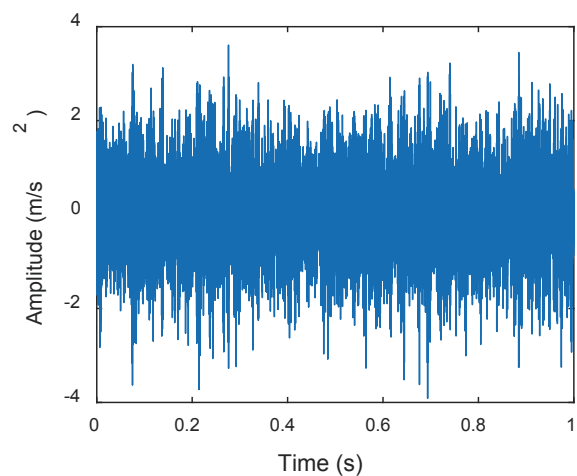
$$f_r = \frac{n}{60}; \text{ Rotational frequency (14)}$$



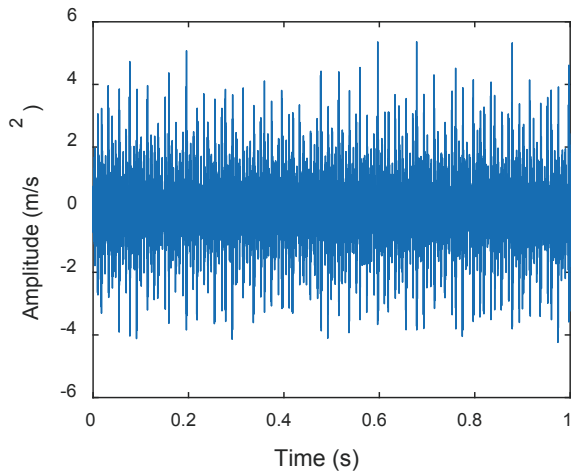
**Figure 2** Healthy state vibration signal. The data of figure 2 to 8 were normalized.



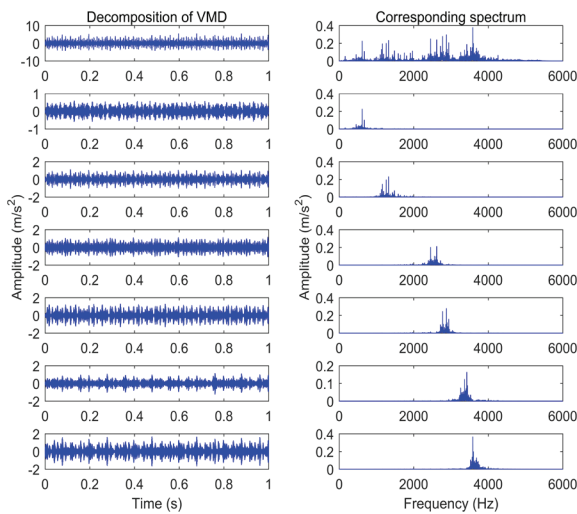
**Figure 3** Outer ring fault vibration signal



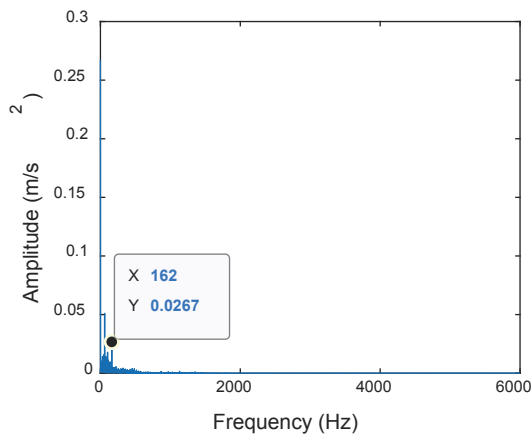
**Figure 4** Bearing element fault vibration signal



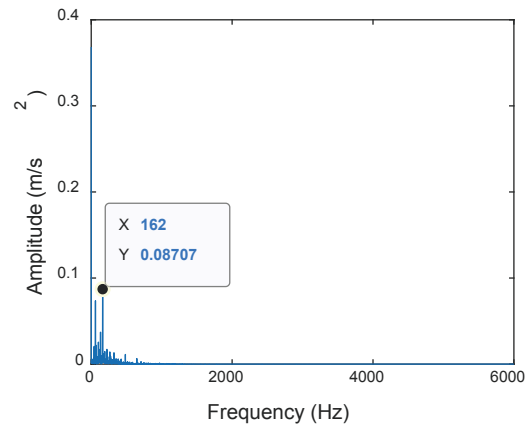
**Figure 5** Inner ring fault vibration signal



**Figure 6** Modal function and spectrum of VMD. The inner ring fault signal composed of 10000 data points was decomposed into six modal components.



**Figure 7** Envelope spectrum of the 1st modal component. Fault frequency is not obvious.



**Figure 8** Envelope spectrum of the 2nd modal component. Fault frequency is obvious.

### 3.3 Time Domain Feature Extraction

The 10 time domain features of bearing signals were extracted: peak  $f_0$ , peak index  $f_1$ , skewness  $f_2$ , kurtosis  $f_3$ , pulse index  $f_4$ , root mean square value  $f_5$ , kurtosis factor  $f_6$ , margin index  $f_7$ , waveform index  $f_8$  and root square amplitude  $f_9$ , and the above features are composed into the vector  $F=[f_0, f_1, \dots, f_9]$ . The early bearing vibration data of CWRU were extracted, and 1024 points were taken as a group, 15 sets of time domain eigenvectors were extracted respectively from 4 kinds of load (speed) and 4 kinds of state types (in which the fault diameter is 0.178mm), and 240 sets of data were divided into 160 training sets and 80 test sets.

### 3.4 Time-Frequency Feature Extraction

In literature [17], the frog jump algorithm was used to optimize the variational modal decomposition by taking the minimum value of envelope entropy as the objective function. The number of modal function  $K=6$  or  $7$  and the coefficient of quadratic penalty  $\alpha=2000$  or  $3000$  were obtained for different fault states.

Same as 3.3, the variational mode decomposition of  $K=6$  and  $\alpha=2000$  was performed on the bearing data with 1024 points as a group. The six modal components obtained by decomposition of the signal are composed into the modal matrix, and the singular value decomposition (SVD) was applied to the modal matrix. 1 set of feature vectors is denoted as  $F=[f_0, f_1, f_2, f_3, f_4, f_5]$ , a total of 240 groups, which were divided into 160 training groups and 80 test groups.

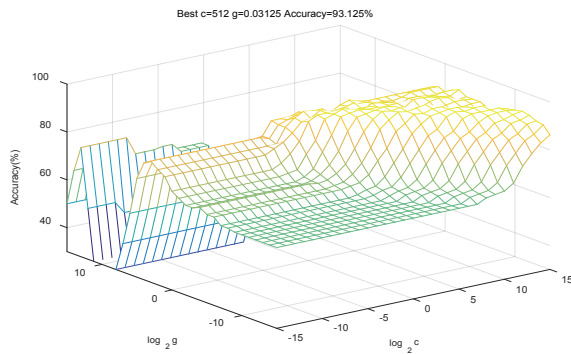
## 4. Fault Identification

### 4.1 Fault State Recognition Based on Time Domain Feature

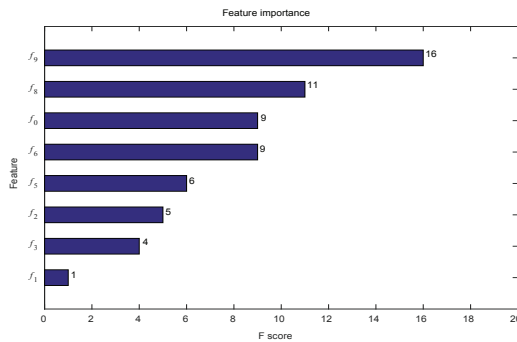
1) Input the obtained time domain feature vectors into the support vector machine after grid optimization for classification [18] (optimization penalty parameter  $c$  and radial basis kernel function parameter  $g$  and  $c, g \in [2^{-15}, 2^{15}]$ ), and the classification accuracy was 93.125%, as shown in fig.9.

2) Input data into xgboost after parameter adjustment (learning rate =0.01, number of trees n=1000), got the accuracy of 91.25%, and got the importance order of time domain features, as shown in fig.10.

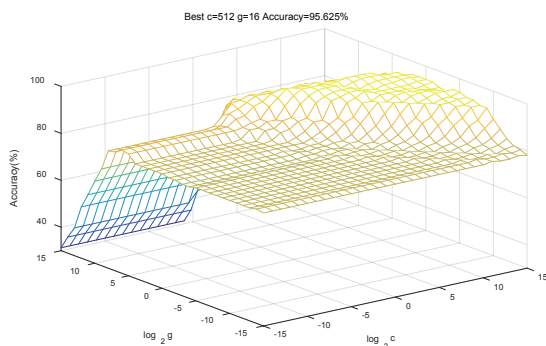
3) Take the top five features, namely peak, root mean square value, kurtosis factor, waveform index and root square amplitude were taken to form the feature vector  $F'=[f_0, f_5, f_6, f_8, f_9]$  and input it into the SVM after parameter adjustment, and the accuracy was 95.625%, as shown in fig.11.



**Figure 9** Time domain feature classification accuracy. In the range of c and g, grid search was carried out to get the maximum accuracy.



**Figure 10** Order of importance of time domain features. A histogram obtained by ranking input time domain features by importance by xgboost.



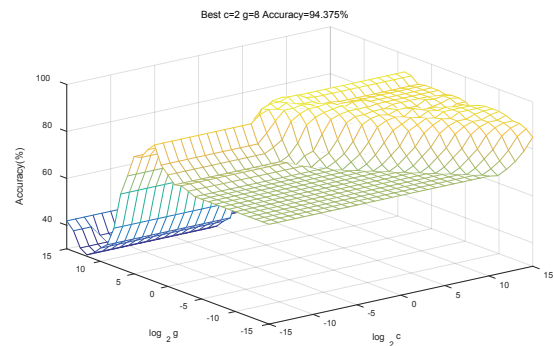
**Figure 11** Classification accuracy after time domain feature extraction. After feature extraction, the accuracy obtained by grid search within the range of c and g.

#### 4.2 Fault State Recognition Based on Time-Frequency Domain Feature

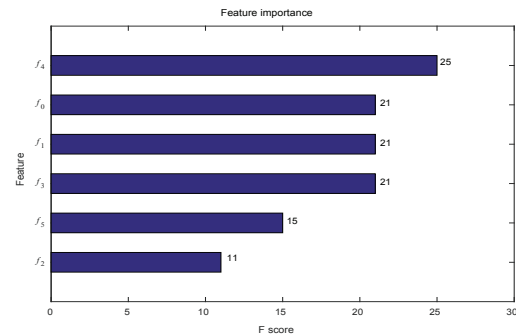
1) Input the time-frequency domain feature vectors into the support vector machine after grid optimization for classification ( $c, g \in [2^{-15}, 2^{15}]$ ), and obtain the classification accuracy was 94.375%, as shown in fig.12.

2) Input data into Xgboost after parameter adjustment (learning rate =0.1, number of trees n=500), got the accuracy was 86.25%, and the importance order of features was obtained, as shown in fig.13.

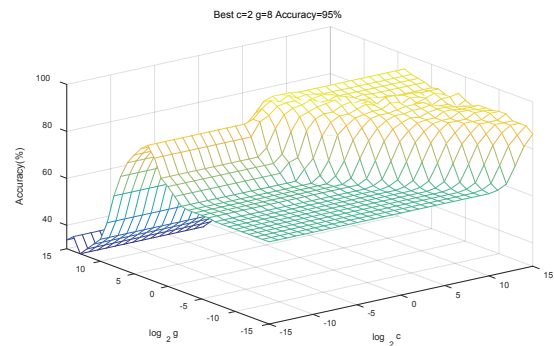
3) Take the top five features features  $f_0, f_1, f_3, f_4, f_5$ , and compose the vector  $F'=[f_0, f_1, f_3, f_4, f_5]$ , and input it into the SVM which was adjusted, and the accuracy was 95%, as shown in fig.14.



**Figure 12** Time-frequency domain feature classification accuracy.



**Figure 13** Order of importance of singular values



**Figure 14** Classification accuracy after time-frequency domain feature extraction

### 4.3 Result Analysis

It can be seen from 4.2 and 4.3 that the Xgboost algorithm can improve the accuracy of fault diagnosis by ranking the time domain and time-frequency domain features respectively and extracting important features for fault diagnosis.

Doing 4 sets of experiments according to the method proposed in this article, and the results are presented in Table.1 and Table.2, it can be seen that the proposed method is more accurate than SVM and Xgboost algorithm.

**Table.1.** Accuracy of time domain feature extraction

Data	SVM (%)	Xgboox (%)	Xgboox+SVM (%)
1	93.125	91.5	95.625
2	83.75	75.625	85.625

**Table.2.** Accuracy of time-frequency domain feature extraction

Data	SVM (%)	Xgboox (%)	Xgboox+SVM (%)
1	94.375	86.25	95
2	87.5	70	88.75

### 5. Conclusion

In view of the influence of noise and other information on signal feature extraction in bearing fault diagnosis, Xgboost algorithm is used in this paper to order the importance of features, the important features in time domain and time frequency domain are extracted and fault diagnosis studied. Through the study, the conclusions are as follows:

1) For fault diagnosis of rolling bearing, some features directly obtained by feature extraction method do not play a role in fault diagnosis. On the contrary, as the dimension of feature vector increases, overfitting and other phenomena will occur. The results of signal analysis show that the accuracy of fault diagnosis can be improved by using Xgboost algorithm to prioritize the features of signals in time domain and time-frequency domain, and by extracting important features to reduce the feature dimension for fault diagnosis.

2)The VMD algorithm optimized by group intelligence algorithm tends to fall into the local optimal solution. The mode matrix composed by VMD is decomposed by SVD to construct the feature vector, and the important features are extracted by Xgboost algorithm, which can improve the accuracy of fault diagnosis.

**Author Contributions:**W.X.G. and W.C. contributed the central idea, analysed the data, and wrote the paper.

**Conflict of Interest:**The authors declare that there is no conflict of interest regarding the publication of this paper.

**Acknowledgments:**This work was supported by The National Natural Science Foundation of China (No.51475086), Fundamental Research Funds for the Central Universities of China (No. N162312001), and CAST-BISEE Foundation (No. CAST-BISEE2019-019).

### References

- [1] WANG Xingang, YAN Mingming. Research on tool change time and dynamic reliability of machining process based on sensitivity analysis. *Journal of Ordnance Equipment, Engineering*, 2017, 38(1), pp.1-6. (in chinese)
- [2] ZHANG Yu, CHEN Jun, WANG Xiaofeng, et al. Application of Xgboost to fault diagnosis of rolling bearings. *Noise and Vibration Control*, 2017, 37(4), pp.166-170. (in chinese)
- [3] WANG Lei, LIU Zhiwen, MIAO Qiang, et al. Time-frequency analysis based on ensemble local mean decomposition and fast kurtogram for Rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 2018, 103, pp.60-75.
- [4] Dragomiretskiy K, Zosso D. Variational mode decomposition. *IEEE. Transactions on Signal Processing*, 2014, 62(3), pp.531-544.
- [5] TANG Guiji, WANG Xiaolong. Parameter optimized variational mode decomposition method with application to incipient fault diagnosis. *Journal of Xi'an Jiaotong University*. 2015, 49(5), pp.73-81. (in chinese)
- [6] WANG Xianbo, YANG Zhixin, YAN Xiaolan. Novel particle swarm optimization based variational mode decomposition method for the fault diagnosis of complex rotating machinery. *Energy*, 2018, 23(1), pp.68-79.
- [7] LIU Jianchang, QUAN He, YU Xia, et al. Rolling bearing fault diagnosis based on parameter optimization VMD and sample entropy. *Acta Automatica Sinica* (in press). (in chinese)
- [8] PAN Jun, ZI Yanyang, CHEN Jinglong, et al. Lifting Net: A novel deep learning network with layer wise feature learning from noisy mechanical data for fault classification. *IEEE. Transactions on Industrial Electronics*, 2018, 65(6), pp.4973-4982.
- [9] FU Yunxiao, JIA Limin, QIN Yong, et al. Rolling bearing fault diagnosis method based on LMD-CM-PCA. *Journal of Vibration, Measurement &Diagnosis*, 2017, 37(2), pp.249-255. (in chinese)
- [10] ZHANG Xining, ZHANG Wenwen, ZHOU Rongtong, et al. Bearing fault diagnosis method based on multiple dimensional scaling and random forest. *Journal of Xi'an Jiaotong University*, 2019, 53(8), pp.1-7. (in chinese)
- [11] Vapnik V. *The nature of statistical learning theory*, 2nd ed. Springer:New York, America; 1995; pp.163-167.
- [12] Zidi S, Moulahi T, Alaya B. Fault detection in wireless sensor networks through SVM classifier. *IEEE. Sensors Journal*, 2018, 18(1), pp.211-244.

- [13] ZHAN Chunfu, WANG Song, WU Yadong, et al. Diabetes risk prediction based on GA\_Xgboost model. Computer Engineering (in press).(in chinese)
- [14] CHEN Tianqi, Guestrin C. Xgboost: A Scalable Tree Boosting System. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM:California, America, 2016, pp.785-794.
- [15] Bearing Data Center Website. Available online:<http://csegroups.case.edu/bearingdatacenter/home>. (accessed on 10.5.2019)
- [16] LI Jimeng, YAO Xifeng, WANG Hui, et al. Periodic impulses extraction based on improved adaptive VMD and sparse code shrinkage denoising and its application in rotating machinery fault Diagnosis. Mechanical Systems and Signal Processing. 2019, 126, pp.568-589.
- [17] MA Hongbin, TONG Qingbin, ZHANG Yanan. Applications of optimization parameters VMD to fault diagnosis of rolling bearings. China Mechanical Engineering, 2018, 29(4), pp.390-397. (in chinese)
- [18] LIU Ruonan, YANG Boyuan, ZIO E, et al. Artificial intelligence for fault diagnosis of rotating machinery:A Review. Mechanical Systems and Signal Processing. 2018, 108, pp.33-47.