

Article

Naive Bayes Bearing Fault Diagnosis Based on Enhanced Independence of Data

Nannan Zhang ^{1,2,3,4}, Lifeng Wu ^{1,2,3,4,*}, Jing Yang ^{1,2,3,4} and Yong Guan ^{1,2,3,4}

¹ College of Information Engineering, Capital Normal University, Beijing 100048, China; zhangnn@cnu.edu.cn (N.Z.); yangjing@cnu.edu.cn (J.Y.); guanyong@cnu.edu.cn (Y.G.)

² Beijing Key Laboratory of Electronic System Reliability Technology, Capital Normal University, Beijing 100048, China

³ Beijing Key Laboratory of Light Industrial Robot and Safety Verification, Capital Normal University, Beijing 100048, China

⁴ Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China

* Correspondence: wulifeng@cnu.edu.cn; Tel.: +86-134-0110-8644

Received: 21 December 2017; Accepted: 1 February 2018; Published: 5 February 2018

Abstract: The bearing is the key component of rotating machinery, and its performance directly determines the reliability and safety of the system. Data-based bearing fault diagnosis has become a research hotspot. Naive Bayes (NB), which is based on independent presumption, is widely used in fault diagnosis. However, the bearing data are not completely independent, which reduces the performance of NB algorithms. In order to solve this problem, we propose a NB bearing fault diagnosis method based on enhanced independence of data. The method deals with data vector from two aspects: the attribute feature and the sample dimension. After processing, the classification limitation of NB is reduced by the independence hypothesis. First, we extract the statistical characteristics of the original signal of the bearings effectively. Then, the Decision Tree algorithm is used to select the important features of the time domain signal, and the low correlation features is selected. Next, the Selective Support Vector Machine (SSVM) is used to prune the dimension data and remove redundant vectors. Finally, we use NB to diagnose the fault with the low correlation data. The experimental results show that the independent enhancement of data is effective for bearing fault diagnosis.

Keywords: Naive Bayes; decision tree; support vector machines; fault diagnosis

1. Introduction

The rolling bearing is the main component of rotating machinery. It carries the entire rotating machinery and equipment operation, and a small fault may have a significant impact on the operation of the entire device. Most of the problems with rotating machines are caused by bearing failure [1]. Therefore, the bearing fault diagnosis is of great significance. After the fault diagnosis of rotating machinery, the machine can be repaired and handled in time, so as to avoid the catastrophic effect caused by mechanical failure [2]. The related contents and techniques of fault diagnosis are introduced in the literature [3–6]. Before the machine failure, the maintenance and treatment of the machine can prevent the probability of failure and reduce the maintenance costs of the machine, as well as avoid casualties caused by equipment failure.

The vibration analysis is the main tool for the diagnosis of rotating machinery [7], and vibration signals analysis has been widely used in the field of fault diagnosis. In this field, the vibration spectrum analysis technique has successfully identified the faults [8–11]. Through the analysis of vibration signals, the state of rotating machinery can be reflected. Sensors can be used to collect vibration signals of

operating machinery, which contains rich information about the working state of machinery [12]. The mechanical health state is determined by analyzing the collected vibration signals. However, the collected vibration signals are chaotic and irregular. Therefore, it is necessary to extract the most representative, reliable and effective features from the acquired vibration signals.

The time domain signal feature of statistical analysis can be used to detect faults, which is mainly to extract feature of data. However, it can only reflect whether rotating machinery and its state are normal, and it can give diagnostic messages but not diagnose the fault, so further fault diagnosis is needed. Nowadays, with the successful application of machine learning methods in various fields, more and more machine learning methods are used in mechanical fault diagnosis. Neural networks, as a typical method of machine learning, has been applied to the field of fault diagnosis. As the most popular classifier, Support Vector Machines (SVM) has achieved some success in the field of fault diagnosis. SVM is a powerful tool for classification, and it also plays a significant role in machine fault diagnosis [13]. Samanta [14] proposed time-domain characteristics of the rotating machine that can be used as the input of artificial neural network (ANN) to verify the effective application in bearing fault diagnosis. Jack et al. [15] put forward SVM for bearing fault diagnosis. Yang et al. [16] proposed that vibration signals can be decomposed to stationary intrinsic mode functions (IMFs), and the input of ANN is the energy features extracted from IMF so as to identify the rolling bearing failure. Al-Raheem et al. [17] proposed a new technique that used genetic algorithm to optimize the application of Laplace wavelet shape and classifier parameters of ANN for bearing faults. In addition, Shuang et al. [18] proposed a fault pattern recognition method based on the principal component analysis (PCA) and SVM. However, the extracted multi-dimensional feature vector contains a large amount of information, with high data redundancy, which results in higher computational costs. Therefore, the high-dimensional characteristics need to be processed. Wu et al. [19] used the Manifold Learning algorithm to reduce the dimension of the high-dimensional features and then the processed are used as the input of wavelet neural network for bearing fault diagnosis. Sugumaran et al. [20] applied Decision Tree to selecting feature, and then carried on the bearing fault diagnosis with the kernel neighborhood fractional multiple support vector machine (MSVM). In another article [21], first, the time domain statistical feature and histogram feature was extracted from time domain signals, then the main feature was selected by the Decision Tree, last SVM and Proximal Support Vector was used for fault diagnostics of roller bearing. In a recent study, Ran et al. [22] proposed a neural network-based method to directly identify the original time series sensor data without feature selection and signal processing for bearing fault diagnosis. In his other article [23], the network is a combination of linear and nonlinear method, and also uses the depth network classifier of the original time series sensor data to diagnose faults.

ANN and SVM have an extensive application in fault diagnosis. However, there are some limitations. For example, the fitting problem and the local extremum can lead to slow operation speed and inaccuracy in ANN training results, respectively [24]. Moreover, SVM has a problem with the speed of testing and training. There are some limitations on multi-class, nonlinear and parameters problem. The training of ANN and SVM is complex, and the cost on training space is high. NB is used not only a small amount of training data, but also its simple structure, fast calculation speed and high accuracy [25,26]. Due to the reliable theoretical basis, comprehensive prior knowledge and the assumption of independent conditions among attributes, NB successfully applied in machine fault diagnosis. Hemantha et al. used the Bayes classifier to diagnose the bearing fault, and verified that NB on fault diagnosis has a good performance [27]. Girish et al. successfully applied NB classifier to the welded joints fault diagnosis [28]. However, the independence assumption of vibration signal of bearing fault is difficult to be realized in actual situations, which limits the algorithm. Therefore, this paper mainly carries on the vector pruning from two aspects of the characteristic attributes and the data dimension. First, Decision Trees are mainly used to select the main feature attributes [29]. Then, the redundancy of dimension vectors is removed by the proposed selective support vector

machine (SSVM). In this way, the redundant data is processed from two aspects, and the limitation of the independence hypothesis on the NB is reduced. Finally, fault diagnosis model is established.

In this paper, NB, which is based on data independence improvement in fault diagnosis, is proposed. The remainder of paper is organized as follows: in Section 2, there is a brief introduction of the NB model. The fault diagnosis based on improved data independence is given by Section 3. In Section 4, the fault diagnosis based on improved data independence is applied to roller bearing diagnosis. Section 5 draws the conclusion of this paper.

2. NB Model

NB is a supervised learning classification method based on probability. NB has received much attention due to its simple classification model and excellent classification performance. The training model is shown in Figure 1.

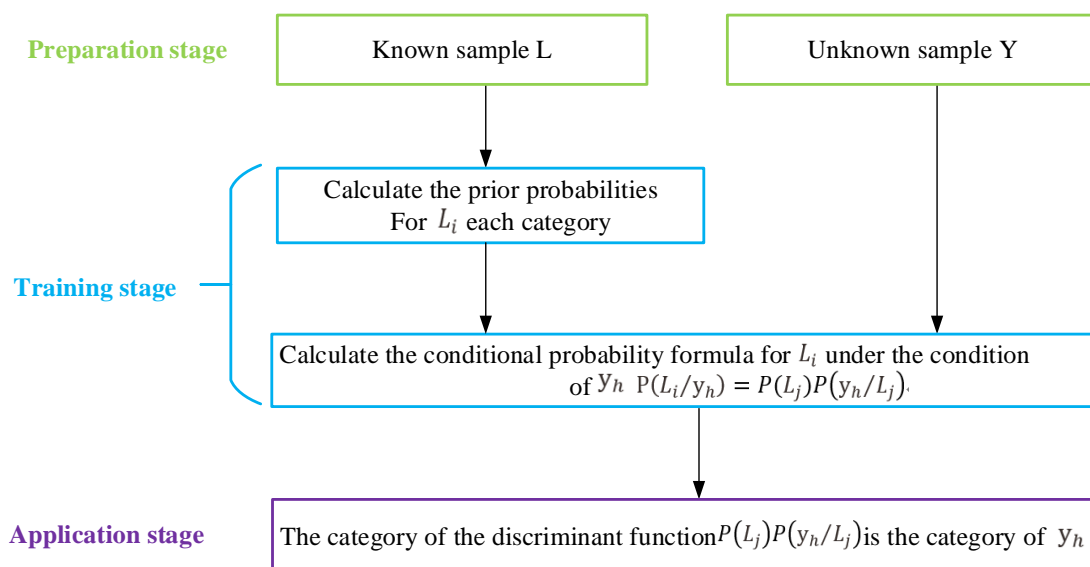


Figure 1. Naive Bayes training model.

(a) Preparatory stage

Suppose there are m categories and categories $L = \{L_1, L_2, \dots, L_m\}$. Each sample has n attributes $At = \{At_1, At_2, \dots, At_n\}$, and each attribute set has d -dimensional feature vector $X = \{X_1, X_2, \dots, X_d\}$.

(b) Training stage

$P(L_i)$ is the prior probability of each category, only related to the ratio of each category to the total category, that is,

$$P(L_i) = \frac{n_i}{n}, 1 \leq i \leq m, \tag{1}$$

where n is the number of known sample, and n_i is the number of i -th categories.

Bayes is a classifier based on the maximum posterior probability. There is an unknown sample classes $Y = \{y_1, y_2, \dots, y_z\}$, and the idea is to calculate the probability of unknown samples in each category. Finally, if the probability of the unknown sample Y is maximum in class L_i , the unknown sample is classified into category L_i . NB is based on the Bayes theorem, and the NB classification method is shown below:

$$P(L_i/y_h) > P(L_j/y_h), 1 \leq i \leq m, 1 \leq j \leq m, 1 \leq h \leq z, i \neq j. \tag{2}$$

According to Bayes's theorem, the probability formula of $P(L_i/y_h)$ can be obtained. The NB is a Bayes theorem based on the independence of the characteristic conditions, so $P(L_i/y_h)$ can be defined as follows:

$$P(L_i/y_h) = \frac{P(y_h/L_i)P(L_i)}{P(y_h)}, \quad (3)$$

$$P(y_h) = \sum_{i=1}^m P(y_h/L_i)P(L_i), \quad (4)$$

where $P(y_h)$ is a constant, and it is only necessary to compute the formula $P(y_h/L_i)P(L_i)$ of Equation (3).

According to the NB classification method, the value of the discriminant function $P(y_h/L_i)P(L_i)$ in each class is calculated for the unknown sample, where $P(L_i)$ is a priori probability of each category, as shown in Equation (1), and where $P(y_h/L_i)$ is the probability of y_h under the condition of L_i . The attribute At_{gi} is continuous property and independent of each other. In general, the attribute variable obeys the Gaussian distribution $At_{gi} \sim N(u_{gi}, \delta_{gi}^2)$ [30]; then, $P(y_h/L_i)$ is defined as follows:

$$P(y_h/L_i) = \frac{1}{\sqrt{2\pi}\delta_{gi}} \exp \left\{ -\frac{(y_h - u_{gi})^2}{2\delta_{gi}^2} \right\}, \quad (5)$$

where u_{gi} and δ_{gi}^2 are mean and variance of samples, respectively, and the formula is as follows:

$$u_{gi} = \frac{\sum_{i=1}^{n_i} X_{ig}}{n_i}, \quad (6)$$

$$\delta_{gi}^2 = \frac{\sum_{i=1}^{n_i} (X_{ig} - u_{gi})^2}{n_i - 1}. \quad (7)$$

From the above Equations (2) and (5)–(7), the posterior probability equation can be obtained:

$$P(L_i/y_h) = P(L_i) \prod_{g=1}^n \frac{1}{\sqrt{2\pi}\delta_{gi}} \exp \left\{ -\frac{(y_h - u_{gi})^2}{2\delta_{gi}^2} \right\}. \quad (8)$$

In the same way:

$$P(L_j/y_h) = P(L_j) \prod_{g=1}^n \frac{1}{\sqrt{2\pi}\delta_{gj}} \exp \left\{ -\frac{(y_h - u_{gj})^2}{2\delta_{gj}^2} \right\}. \quad (9)$$

(c) Application stage

According to the Equation (2), if $P(L_i)P(L_i/y_h) > P(L_j)P(L_j/y_h)$, the unknown sample is judged as class i ; otherwise, it is judged as j .

3. NB Fault Diagnosis Model Based on Enhanced Independence of Data

3.1. Fault Diagnosis Model

In order to improve the classification effect of NB, this paper enhances the independence between data from two aspects of attribute characteristics and data dimension. The proposed fault diagnosis model is shown in Figure 2. The fault diagnosis model includes three parts: signal acquisition, signal processing and fault diagnosis.

- Signal acquisition: Acceleration sensor is used to obtain vibration signals of rolling bearings.
- Signal processing: The original vibration signal of the rolling bearing obtained from the sensor contains a large amount of noise, so it is necessary to process the data to obtain valid data signals.

Firstly, feature extraction is performed on the original signal acquired by using the time-domain signal method. Then, the Decision Tree is used to select the main feature attributes from the feature attributes. The data are processed from two directions of feature attribute and data dimension, so that the data with strong independence can be obtained, which is beneficial to the fault diagnosis of the bearing.

- Fault diagnosis: After the data is processed, we obtain data with low redundancy. Thus, the impact of data independence assumption on NB model is reduced, and the fault diagnosis can be made effectively.

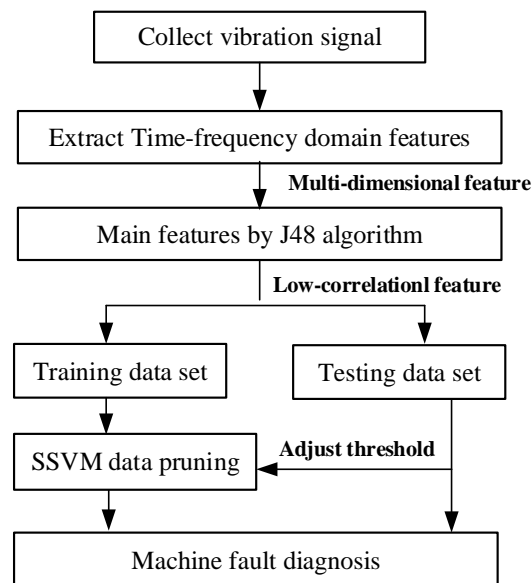


Figure 2. Fault diagnosis model based on the enhanced independence of data.

3.2. Feature Selection Using Decision Tree

The Decision Tree is a tree structure, which is mainly composed of nodes and branches, and the nodes contain leaf nodes and intermediate nodes. The intermediate nodes are used to represent a feature, and leaf nodes are used to represent a class label. The Decision Tree can be used for feature selection [29]. The attributes that appear in the Decision Tree nodes provide important information to promote classification. The J48 algorithm is mainly used to construct Decision Tree. Therefore, we construct a Decision Tree using J48 algorithm. Then, we find the characteristic attribute corresponding to the middle node of the decision tree, and remove the feature attribute that without important information. The following describes the J48 algorithm for feature extraction:

- The acquired data is used as the input of the algorithm, and the output is the node of the Decision Tree.
- The output Decision Tree nodes are divided into leaf nodes and intermediate nodes. The leaf node represents the classification, the intermediate node represents the decision attribute, and the branch represents the condition that the next decision data comes from the previous decision attribute.
- The Decision Tree is used to find feature attributes from top to bottom until all nodes become leaf nodes.
- Finding the criteria of decision attributes: the information gain of each feature is calculated and the maximum information gain is chosen as the intermediate node of the Decision Tree.

Information gain is used to determine how to select the most appropriate features from a number of attributes. Information gain is mainly determined by the information entropy. Information gain of

attribute At for the data set is: entropy of all attribute information minus the entropy of split attributes. The At is a continuous attribute based on Gaussian distribution, so information entropy properties of At is defined as follows:

$$Gain(At) = Info(L) - info_{At}(L), \quad (10)$$

$$Info(L) = - \sum_{j=1}^m P(j/L) \log P(j/L), \quad (11)$$

$$Info_{At}(L_j) = \sum_{j=1}^m \frac{L_j}{L} \frac{\text{Log} 2\pi e \delta_{ij}^2}{2}. \quad (12)$$

$Gain(At)$ is the information gain of the attribute At , $Info(L)$ is the undivided information entropy, and $info_{At}(L)$ is the information entropy At after splitting. The variance x is given by the Formula (6), and m is the number of classifications, and L_j is a subset of data set L .

3.3. SSVM

SVM is a traditional classification method for two categories. In this paper, an optimal classification hyperplane is constructed in the sample set, and two classes of samples are separated from each other on the hyperplane. Generally, in the case of too much data, SVM can not completely classify the two kinds of data into both sides of the hyperplane. Thus, we propose an SSVM algorithm to remove the spatial redundancy problem of the vector.

SSVM data processing is divided into several steps, as shown in Figure 3.

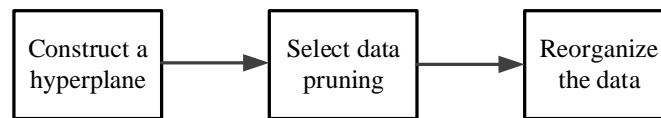


Figure 3. Selective Support Vector Machine data processing flow chart.

Step 1: Constructing the optimal hyperplane of data.

In most cases, SVM is targeted at two types of problems [31]. The data set (X, Y) is divided into training set and test set. The training set is $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. if X_i is the first class, $Y_i = 1$. if X_i is second class, $Y_i = -1$. As shown in Figure 4, hyperplane $H(X)$ separates the two-class data on both sides.

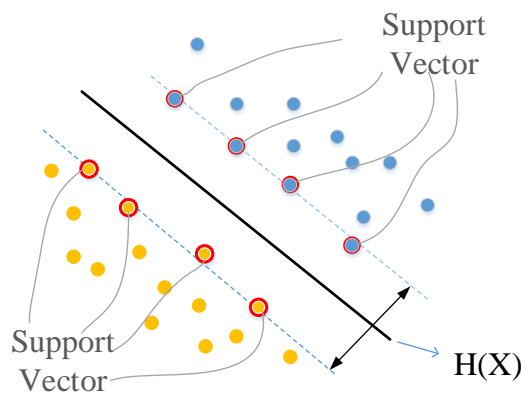


Figure 4. Two categories of Support Vector Machine.

The hyperplane $H(X)$ equation is given as in Equation (13) [32]:

$$w^T K(X) + b = 0. \quad (13)$$

The function $K(X)$ is a kernel function, which maps the low-dimensional space to the high-dimensional space, and avoids the fact that the data cannot be separated in the low-dimensional space, where w is a vector, b is constant, and their values can be obtained by the optimization of the following Equation [31]:

$$\min : \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i, \|w\| = w^T w, \quad (14)$$

$$s.t. \quad y_i(w^T K(X_i) + b) \geq 1 - \xi_i, \xi_i \geq 0. \quad (15)$$

Parameter C is mainly used to adjust training error. ξ_i is a slack variable [33]. After the solution of the parameter, the optimal hyperplane $H(X)$ is obtained [31]:

$$H(X) = \text{sgn}\left(\sum_{i \in SV} y_i a_i K(x_i, x) + b\right), \quad (16)$$

where S is the support vector for the DataSet (X, Y) , where sgn is a symbolic function that mainly returns the positive and negative of the parameter. $K(x_i, x)$ is a kernel function, and there are many kinds of kernel functions. The Gaussian kernel function is better in the application, so the Gaussian kernel function is used in this paper:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right). \quad (17)$$

Step 2: Using the constructed hyperplane to select the data and remove the redundancy.

Firstly, a suitable threshold is selected, and the hyperplane $K(X)$ is used to test the data. When the test result does not reach the threshold, this data is chosen to be pruned.

Then, find the hyperplane boundary support vector.

Finally, find the point closest to each support vector, and judge if the closest distance is consistent with the classification of the vector; then, keep it, or otherwise delete it.

This article uses the Euclidean distance to measure the distance between two points. For high-dimensional data, the distance between two points is the distance of two vectors, for example, $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$, X and Y distance $D(X, Y)$ is written as:

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, 1 \leq i \leq n. \quad (18)$$

Step 3: Reorganizing processing data, and obtaining new data.

SVM is mainly used for two types of data. This article mainly uses multiple categories of data. First of all, the data in multiple categories were put into pairs, respectively. Then, two kinds of data are pruned with SSVM. The data processing is divided into the following steps:

- (1) Construct hyperplane for training.
- (2) Test the data with a trained hyperplane.
- (3) Set the appropriate threshold to find out the classification of the training data and training results below the threshold.
- (4) Finding the nearest neighbor of each support vector form data obtained in step (3), calculating the distance between the support vectors and the data points, and setting the distance between the points to itself be infinity.
- (5) Find the nearest vector point of each support vector.

- (6) Determine whether the support vector is consistent with its corresponding nearest neighbor vector classification result, and mark it as 0 if inconsistent.
- (7) Remove the data marked as 0 in the data.
- (8) Reorganize data to get new data.

According to the description of the SSVM, the SSVM pruning algorithm is the most important part of the SSVM. The details of SSVM pruning algorithm are shown in Algorithm 1.

Algorithm 1 SSVM pruning algorithm.

Input:

The selected training sample $\langle X, Y \rangle, X = (X_1, X_2, \dots, X_n)$;

Output:

Trimmed sample $\langle X_1, Y_1 \rangle$

```

1:  Begin
2:  Obtain support vector  $\langle Z, H \rangle$  by SVM,  $Z = (Z_1, Z_2, \dots, Z_n)$ 
3:  for  $i:=1$  to  $n$  do
4:    for  $j:=1$  to  $m$  do
5:      Calculate the distance  $D(Z_i, X_j)$  between  $Z$  and  $X$  by Equation (18), When  $Z_i$  is the
      same as  $X_j$ , define the distance  $D$  as infinite.
6:    end
7:    Find the nearest dimension vector  $X_j$  between  $Z_i$  and  $X$ 
8:    Judge whether  $H_i$  and  $Y_j$  are the same, if not, let  $Y_j = 0$ 
9:  end
10: Delete the sample data of the  $Y=0$ 
11: return  $\langle X_1, X_2 \rangle$ 
12: end

```

4. Experiment and Analysis

4.1. Bearing Data Preprocessing

The data in this article is from bearing fault signals provided by the Case Western Reserve University (CWRU) laboratories [34]. The experimental platform is shown in Figure 5. The experimental platform consists of a torque tachometer, a 1.5 KW motor and a dynamometer. The experimental data uses the acceleration signals collected by the acceleration sensors. The sensor is fixed to the position of the driving end and the fan end of the motor shell at 12 o'clock with the magnetic base, and the vibration signal is collected through the recorder. The type of bearing used in the test is SKF6205-2RS deep groove ball bearing. The sampling frequency of the experiment is 12 KHz, the speed is 1797 rpm, and the main data is collected from normal vibration signal and the fault vibration signal.

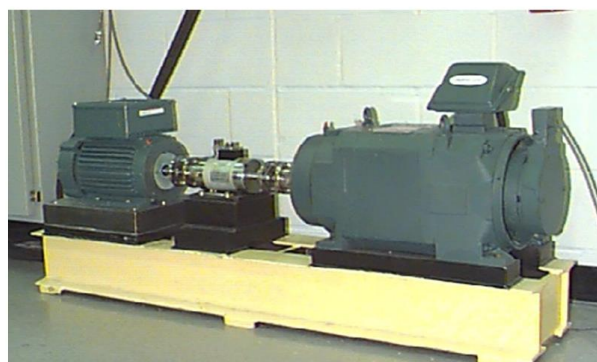


Figure 5. Experimental diagram of experimental platform for rolling bearing fault.

In this paper, the normal vibration signals and fault signals of bearings are analyzed, and the samples of each type of signals are at least 12,100. The main samples of this paper are those samples with no load and the 0.021 (inches) radius fault. Table 1 describes a normal bearing signal and five kinds of fault bearing signals used in this paper. Six kinds of bearing data are described in Figure 6.

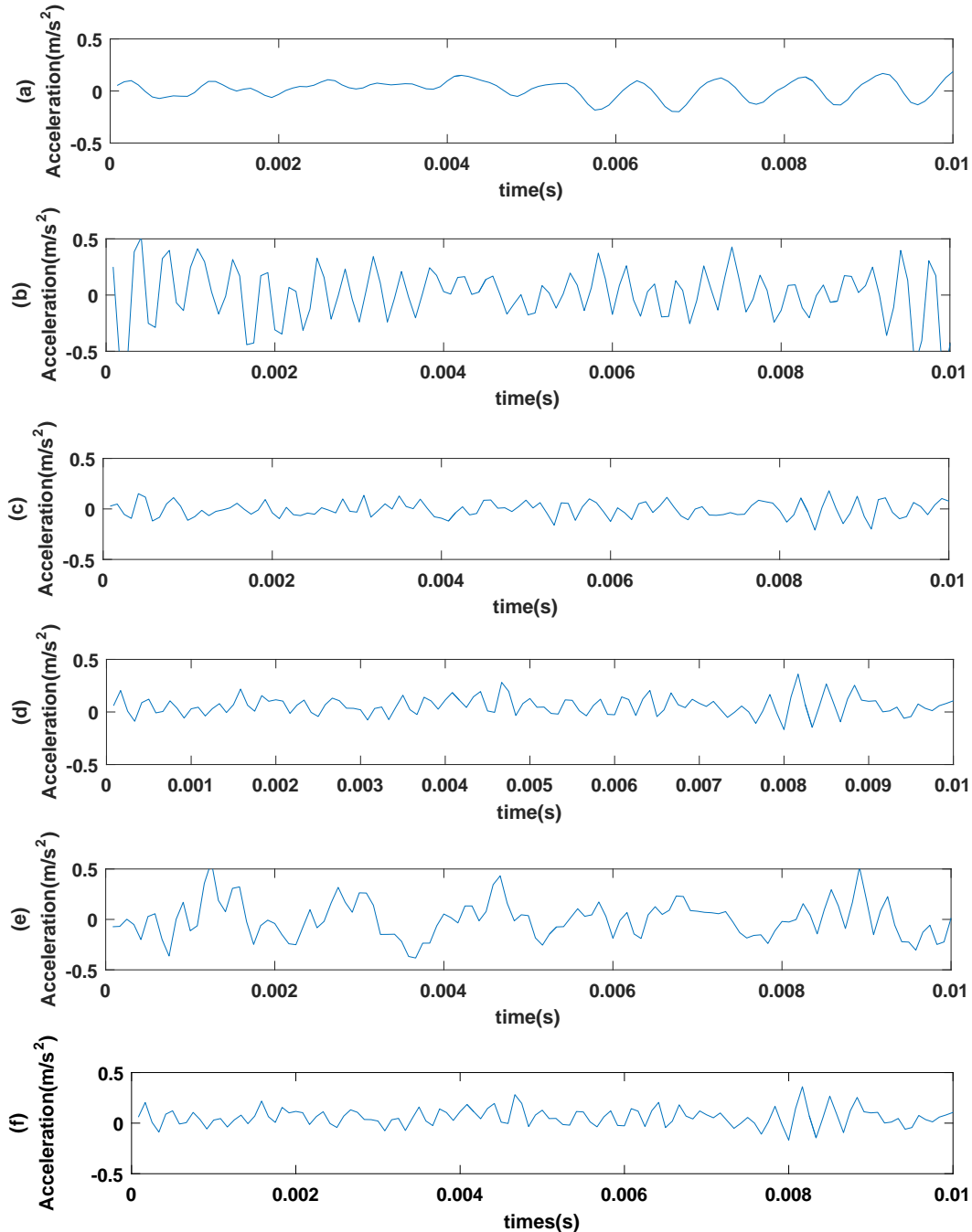


Figure 6. The time domain waveform of rolling bearings is shown in the figure. The x -axis is the time unit of the second and y -axis is the driving end bearing accelerator data. (a) normal bearing signal waveform; (b) inner fault signal waveform; (c) roller fault signal waveform; (d) outer fault signal waveform at center @6:00; (e) outer ring fault signal at orthogonal @3:00; (f) outer fault signal waveform at opposite @12:00.

Table 1. Description of CWRU dataset.

Data Type	Motor Load (HP)	Fault Diameter (Inches)	Label
Normal	0	0	1
Inner race	0	0.021	2
Ball	0	0.021	3
Out race fault at center @6:00	0	0.021	4
Out race fault at orthogonal @3:00	0	0.021	5
Out race fault at opposite @12:00	0	0.021	6

4.2. Application of Improved Algorithm in Bearing Fault

Fault diagnosis model is constructed according to Figure 2. This paper chooses West University of rolling bearing samples and the numbers of each state are at least 121,200. Test data and training data account for half of the total data. The detailed description of various bearing States is shown in Table 2.

Table 2. Description of the data sets.

Data Type	The Number of Training	The Number of Testing	Label
Normal	121	121	1
Inner race	121	121	2
Ball	121	121	3
Out race fault at center @6:00	121	121	4
Out race fault at orthogonal @3:00	121	121	5
Out race fault at opposite @12:00	121	121	6

In this paper, the vibration signal is mainly processed from three aspects.

First, the feature extraction is performed by the time domain method.

The statistical characteristics of signal vibration amplitude will change with the location and the size of the fault. The time domain waveform is dynamically transformed over time. The amplitude of the vibration signal can reflect the characteristic information of the signal intuitively. The time domain waveform information can be used to diagnose the state of the bearing by analyzing the amplitude, shape and other characteristics of the waveform. The time domain characteristic parameters are different due to different fault types and different fault degree. Generally speaking, the time domain feature provides the global characteristics of bearing state, and can effectively extract the bearing fault feature.

In the actual situation, there is various information of bearing fault, and a faults are often accompanied with other faults, such as bearing deformation, corrosion and so on. In order to diagnose the fault more effectively, we need to extract the feature of bearing fault data. In this paper, 17-time domain extraction methods are used to extract the features of the signal.

In Table 3, $X(n)$ is the representative of the signal sample $n = 1, 2, \dots, m$, and m represents the number of samples. Seventeen time domain feature attributes is: T_1 the average value, T_2 absolute mean, T_3 effective value, T_4 average power, T_5 square amplitude, T_6 peak, T_7 peak-to-peak, T_8 variance, T_9 standard deviation, T_{10} skewness, T_{11} kurtosis, T_{12} waveform, T_{13} Crest index, T_{14} impluse index, T_{15} margin index, T_{16} skewness index and T_{17} kurtosis index.

Second, the main feature selection of feature extraction data is made by the Decision Tree.

The main description of the J48 algorithm is given in Chapter 3, and the output tree structure shown in Figure 7. It can be seen from the diagram that the main characteristics of bearing data are T_1 , T_5 , T_{12} and T_{17} .

Table 3. Time domain analysis of bearing fault data.

Number	Characteristic Equation	Number	Characteristic Equation
1	$T_1 = \frac{\sum_{n=1}^d X(n)}{d}$	2	$T_2 = \frac{\sum_{n=1}^d X(n) }{d}$
3	$T_3 = \sqrt{\frac{\sum_{n=1}^d (X(n))^2}{d}}$	4	$T_4 = \frac{\sum_{n=1}^d (X(n))^2}{d}$
5	$T_5 = (\frac{\sum_{n=1}^d \sqrt{ X(n) }}{d})^2$	6	$T_6 = \max X(n) $
7	$T_7 = \max(X(n)) - \min(X(n))$	8	$T_8 = \frac{\sum_{n=1}^d (X(n) - T_1)^2}{d-1}$
9	$T_8 = \sqrt{\frac{\sum_{n=1}^d (X(n) - T_1)^2}{d-1}}$	10	$T_{10} = \frac{\sum_{n=1}^d (X(n))^3}{d}$
11	$T_{11} = \frac{\sum_{n=1}^d (X(n))^4}{d}$	12	$T_{12} = \frac{T_3}{T_2}$
13	$T_{13} = \frac{T_6}{T_3}$	14	$T_{14} = \frac{T_6}{T_2}$
15	$T_{15} = \frac{T_6}{T_5}$	16	$T_{16} = \frac{T_{10}}{T_3}$
17	$T_{17} = \frac{T_{11}}{T_9^4}$		

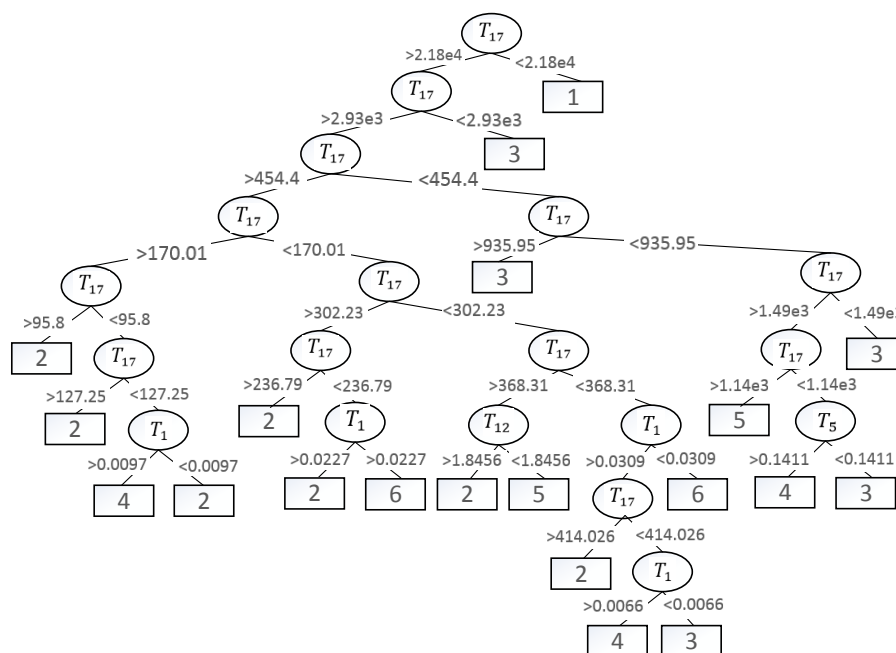


Figure 7. A part of the Decision Tree.

The 17 characteristic attributes obtained by feature extraction are interrelated with each other, which leads to data redundancy. The attributes with low correlation are obtained by extracting the main features with J48 so that the independence of data can be enhanced.

The description and significance of these four main time-domain features are as follows:

- average value (T_1): T_1 is mainly used to reflect the trend of the bearing fault signal,
- square amplitude (T_5): T_5 is mainly used to describe the energy of signals,
- waveform index (T_{12}): T_{12} is sensitive to fault signals with stable waveform,
- kurtosis index T_{17} : kurtosis is sensitive to bearing defects and can reliably reflect the state of rolling bearings. It is not easy to be affected by temperature, speed, etc. and comprehensive analysis of kurtosis, peak factor, and effective value.

In Figure 7, the intermediate node represents the attribute of the decision with an ellipse, and the leaf node represents the classification result with a rectangle. The data between nodes are the classification condition. The graph is a part of the Decision Tree. Class label is a class with the highest probability in classification result when it has little effect on feature selection.

Third, the main feature of extraction is pruned with SSVM.

The J48 algorithm is mainly used to extract attribute vector so that the connection between data is reduced and the independence between data is enhanced. This paper mainly uses SSVM as mentioned above to reduce the similar attributes on the data dimension. The more similar the attribute is, the more redundant it would be. The data redundancy between the pruned data will be reduced so that the independence of the data dimension can be enhanced.

SSVM is used to select the appropriate data for pruning. When the data is removed excessively or removed too little, the classification result will be affected. Therefore, it is very important to choose the appropriate threshold. The threshold in this article is the accuracy rate of test data tested by SVM. When the accuracy is greater than a certain value, we think that these kinds of data are not redundant, so we do not prune it. Therefore, the classification data, which is below the threshold, is selected, and then remove the nearest neighbor inconsistent data. Table 4 shows the selected data corresponding to the pruning data and the pruned training data set, and Figure 8 is the test accuracy of the bearing data corresponding to the selection threshold. From Table 4 and Figure 8, it can be concluded that the data trimming is too small to make the classification effect not obvious, and too much data pruning will result in important data loss. It can be seen from Figure 8 that, when the threshold is 0.9, the corresponding accuracy is the highest than others. Therefore, the training data with a threshold below 0.9 is selected for SSVM pruning. Only in this way can the fault diagnosis be performed effectively.

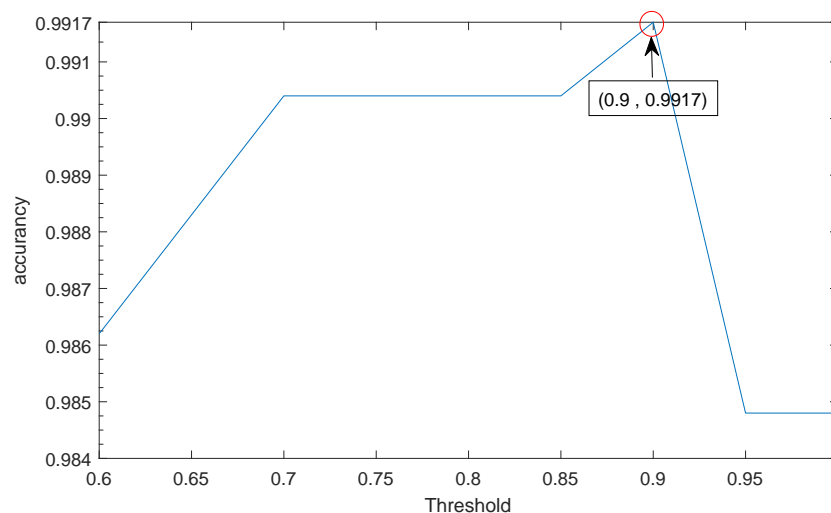


Figure 8. The accuracy of the data corresponding to the threshold.

Table 4. The corresponding threshold data.

Threshold (Training Accuracy)	1	0.95	0.90	0.85	0.80	0.60
The number of pruning	190	187	179	155	102	0
The number of training	536	539	547	571	624	726

After processing, the vibration data from the three aspects above, the redundant data is removed from the feature vector and the dimension vector, respectively. Figure 9 shows the three-dimensional data of time domain feature extraction, three-dimensional data after J48 select feature, and three-dimensional data after J48 and SSVM trimming. The axes x , y , and z in Figure 9 are dimensional features. Among them, Figure 9a selects three dimensions of mean, absolute mean and effective value, Figure 9b,c select three dimensions of mean, waveform index and kurtosis index. It can be seen from Figure 9 that each class of data has obvious overlap in Figure 9a, the overlap ratio of each

kind of data in Figure 9b is obviously lower than of Figure 9a, and Figure 9c obviously separates each type of category data. Therefore, it is shown from Figure 9 that the redundancy between the processed data is greatly reduced, so that the correlation between the data is reduced, and the influence of NB independence assumption on the fault diagnosis is finally reduced.

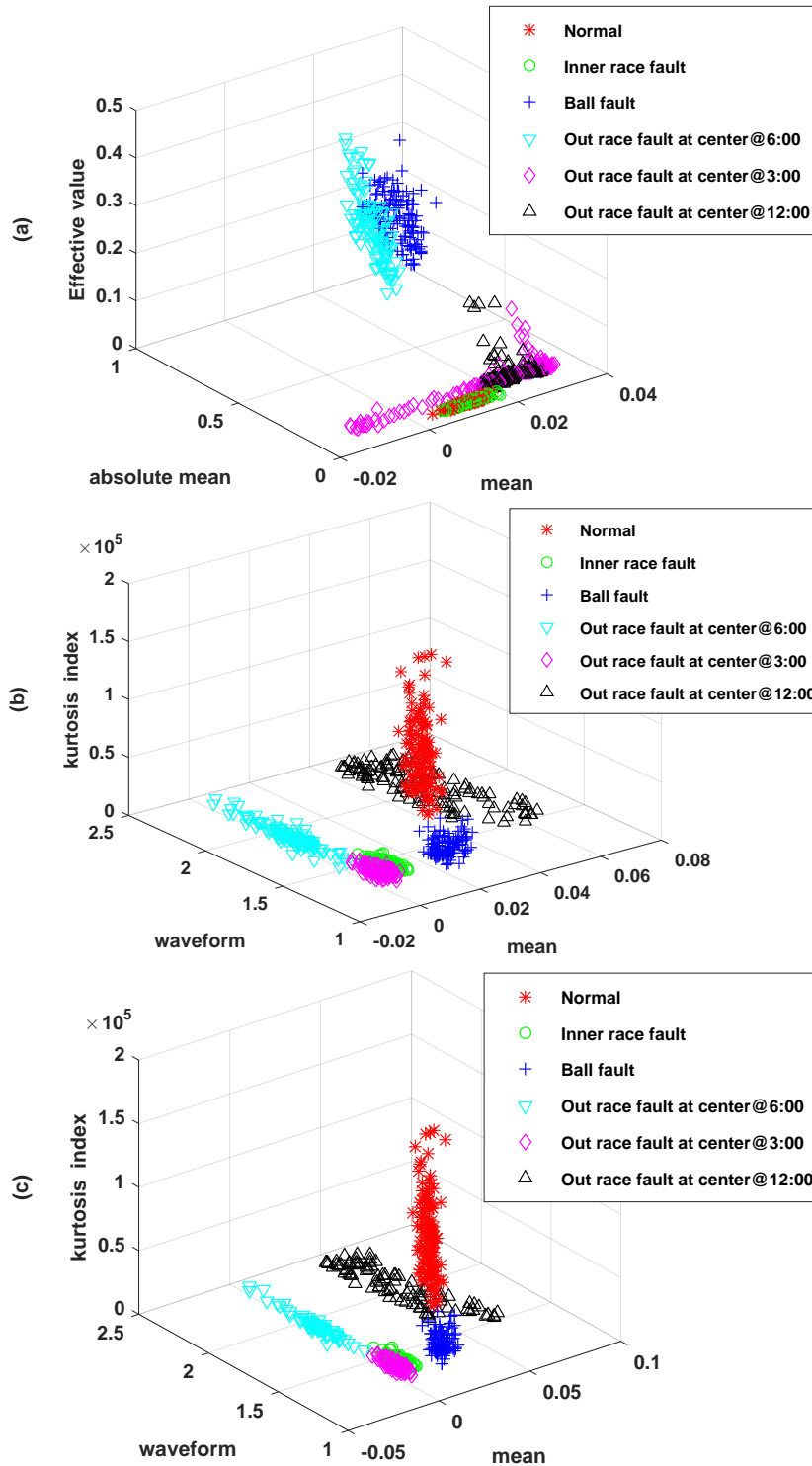


Figure 9. Bearing data description. (a) the original signal time domain feature extraction fault three-dimensional; (b) J48 select the characteristics of the three-dimensional fault data diagram; (c) the three-dimensional fault data diagram after J48 and SSVM pruning.

The processing bearing fault data correlation is low, which reduces the limitation of the independence assumption on NB fault diagnosis. Table 5 is the confusion matrix of NB fault diagnosis for the processed data, and Table 6 is a confusion matrix for bearing fault diagnosis using an NB model without redundant vibration data. As can be seen from the table, the model has been improved for each category after redundancy removal.

Table 5. Confusion matrix of the processing bearing fault data on test sets.

Actual Classes	Predicted Classes					
	1	2	3	4	5	6
1	121	0	0	0	0	0
2	0	121	0	0	0	0
3	0	0	121	0	0	0
4	0	0	0	121	0	0
5	0	0	0	0	121	0
6	0	0	0	0	0	121

Table 6. Confusion matrix of NB on test sets.

Actual Classes	Predicted Classes					
	1	2	3	4	5	6
1	120	0	1	0	0	0
2	0	121	0	0	0	0
3	0	0	117	0	0	4
4	0	1	0	117	2	1
5	0	1	0	3	117	0
6	0	0	0	0	0	121

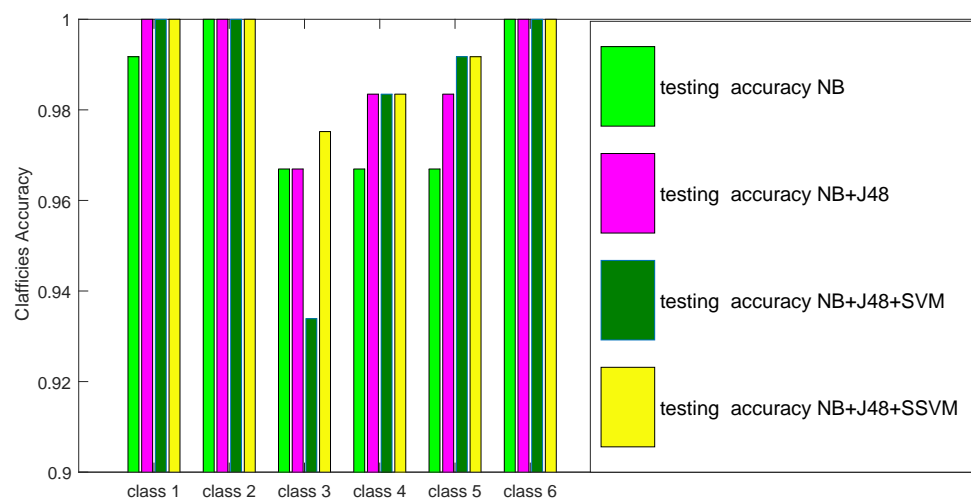


Figure 10. Testing accuracy comparison of each condition in the experiment.

In order to verify the validity of this algorithm in bearing data, the data simulation is carried out by MATLAB (Version 8.6, The MathWorks, MA, USA). Figure 10 and Table 7 are bearing fault diagnosis results. In Figure 10 and Table 7, the meaning of NB+J48+SVM is that first data is selected by J48, then the data after feature selection is pruned by SVM and the fault diagnosis of NB is finally carried out. Compared with other experimental results, the bearing fault diagnosis experimental results on JSSVM-NB is better than removing the data redundancy by feature vector and data vector. Compared with other experiments, the accuracy of the fault diagnosis model is 99.17%. Table 8 shows the comparison of results of about JSSVM-NB and reference [35], which have the same data for bearing

fault diagnosis. It can be seen from Tables 7 and 8 that the JSSVM-NB model is effective for rolling bearing fault diagnosis.

Table 7. The corresponding threshold data.

Methods	Accuracies
NB	98.21%
NB + J48 + SVM	98.48%
NB + J48 + SSVM (JSSVM-NB)	99.17%

Table 8. The comparison results in bearing fault diagnosis.

State	JSSVM-NB	Reference [35]
Normal	100%	98.31%
Inner race	100%	97.73%
Ball	97.5%	95.04%
Out race fault at center (@6:00, @3:00 and @12:00)	99.17%	98.02%

5. Conclusions

In this paper, in order to improve the independence assumption, the bearing data processing is carried out from two aspects of the attribute vector and the dimension vector, and the bearing data with higher data independence is obtained for the bearing fault diagnosis of the NB. NB is based on the conditional independence hypothesis of Bayes. However, in the actual case, it is difficult for the bearing data vector to achieve independence. Therefore, the redundancy is removed from the feature attribute vector and dimension of bearing data in this paper, so that the connection between data is reduced and the bearing condition monitoring on NB be enhanced. It be seen from the simulation results. The NB improved the data independence has realized the fault diagnosis of the different parts of rolling bearing, and can be applied to the other fault diagnosis of the industrial.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (61702348, 61772351, 61602326), the National Key Technology Research and Development Program (2015BAF13B01, the National Key R&D Plan (2017YFB1303000, 2017YFB1302800), and the Project of the Beijing Municipal Science & Technology Commission (LJ201607). The work is also supported by the Youth Innovative Research Team of Capital Normal University.

Author Contributions: Nannan Zhang and Lifeng Wu conceived and designed the experiments; Jing Yang performed the experiments; Yong Guan analyzed the data; Lifeng Wu contributed analysis tools; Nannan Zhang wrote the paper. All authors contributed to discussing and revising the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Jacobs, W.; Malag, M.; Boonen, R.; Moens, D.; Sas, P. Analysis of bearing damage using a multibody model and a test rig for validation purposes. *Struct. Health Monit.* **2011**, *14*, 971–978.
- Li, W.; Mechefske, C.K. Detection of induction motor faults: A comparison of stator current, vibration and acoustic methods. *J. Vib. Control* **2006**, *12*, 165–188.
- Ding, S.X.; Yang, Y.; Zhang, Y.; Li, L. Data-driven realizations of kernel and image representations and their application to fault detection and control system design. *Automatica* **2014**, *50*, 2615–2623.
- Chadli, M.; Abdo, A.; Ding, S.X. H-/H ∞ fault detection filter design for discrete-time Takagi-Sugeno fuzzy system. *Automatica* **2013**, *49*, 1996–2005.
- Chibani, A.; Chadli, M.; Peng, S.; Braiek, N.B. Fuzzy Fault Detection Filter Design for T-S Fuzzy Systems in Finite Frequency Domain. *IEEE Trans. Fuzzy Syst.* **2016**, *25*, 1051–1061.
- Youssef, T.; Chadli, M.; Karimi, H.R.; Wang, R. Actuator and sensor faults estimation based on proportional integral observer for TS fuzzy model. *J. Frankl. Inst.* **2016**, *354*, 2524–2542.

7. Paya, B.; Esat, I.I.; Badi, M.N.M. Artificial Neural Network Based Fault Diagnostics of Rotating Machinery Using Wavelet Transforms as a Preprocessor. *Mech. Syst. Signal Process.* **1997**, *11*, 751–765.
8. Lynagh, N.; Rahnejat, H.; Ebrahimi, M.; Aini, R. Bearing induced vibration in precision high speed routing spindles. *Int. J. Mach. Tools Manuf.* **2000**, *40*, 561–577.
9. Wardle, F.P. Vibration Forces Produced by Waviness of the Rolling Surfaces of Thrust Loaded Ball Bearings Part 1: Theory. *Arch. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **1988**, *202*, 305–312.
10. Mevel, B.; Guyader, J.L. Routes to Chaos in Ball Bearings. *J. Sound Vib.* **2007**, *162*, 471–487.
11. Vafaei, S.; Rahnejat, H. Indicated repeatable runout with wavelet decomposition (IRR-WD) for effective determination of bearing-induced vibration. *J. Sound Vib.* **2003**, *260*, 67–82.
12. Lei, Y.; Lin, J.; He, Z.; Zi, Y. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.* **2011**, *25*, 1738–1749.
13. Wang, T.; Qi, J.; Xu, H.; Wang, Y.; Liu, L.; Gao, D. Fault diagnosis method based on FFT-RPCA-SVM for Cascaded-Multilevel Inverter. *ISA Trans.* **2016**, *60*, 156–163.
14. Samanta, B.; Al-Balushi, K.R. Artificial Neural Network Based Fault Diagnostics of Rolling Element Bearings Using Time-Domain Features. *Mech. Syst. Signal Process.* **2003**, *17*, 317–328.
15. Jack, L.B.; Nandi, A.K. Support vector machines for detection and characterization of rolling element bearing faults. *Arch. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2001**, *215*, 1065–1074.
16. Yang, Y.; Yu, D.; Cheng, J. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *J. Sound Vib.* **2006**, *294*, 269–277.
17. Al-Raheem, K.F.; Roy, A.; Ramachandran, K.P.; Harrison, D.K.; Grainger, S. Application of the Laplace-Wavelet Combined with ANN for Rolling Bearing Fault Diagnosis. *J. Vib. Acoust.* **2008**, *130*, 3077–3100.
18. Shuang, L.; Meng, L. Bearing Fault Diagnosis Based on PCA and SVM. In Proceedings of the International Conference on Mechatronics and Automation, Harbin, China, 5–8 August 2007; pp. 3503–3507.
19. Wu, L.; Yao, B.; Peng, Z.; Guan, Y. Fault Diagnosis of Roller Bearings Based on a Wavelet Neural Network and Manifold Learning. *Appl. Sci.* **2017**, *7*, 158, doi:10.3390/app7020158.
20. Sugumaran, V.; Sabareesh, G.R.; Ramachandran, K.I. Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. *Expert Syst. Appl.* **2008**, *34*, 3090–3098.
21. Sugumaran, V.; Ramachandran, K.I. Effect of number of features on classification of roller bearing faults using SVM and PSVM. *Expert Syst. Appl.* **2011**, *38*, 4088–4096.
22. Zhang, R.; Peng, Z.; Wu, L.; Yao, B.; Guan, Y. Fault Diagnosis from Raw Sensor Data Using Deep Neural Networks Considering Temporal Coherence. *Sensors* **2017**, *17*, 549.
23. Zhang, R.; Wu, L.; Fu, X.; Yao, B. Classification of bearing data based on deep belief networks. In Proceedings of the Prognostics and System Health Management Conference, Chengdu, China, 19–21 October 2017; pp. 1–6.
24. Mohamed, E.A.; Abdelaziz, A.Y.; Mostafa, A.S. A neural network-based scheme for fault diagnosis of power transformers. *Electr. Power Syst. Res.* **2005**, *75*, 29–39.
25. Sharma, R.K.; Sugumaran, V.; Kumar, H.; Amarnath, M. A comparative study of NB classifier and Bayes net classifier for fault diagnosis of roller bearing using sound signal. *Decis. Support Syst.* **2015**, *1*, 115.
26. Mccallum, A.; Nigam, K. A Comparison of Event Models for NB Text Classification. In Proceedings of the AAAI-98 Workshop on Learning for Text Categorization, Madison, WI, USA, 26–27 July 1998; Volume 62, pp. 41–48.
27. Kumar, H.; Ranjit Kumar, T.A.; Amarnath, M.; Sugumaran, V. Fault Diagnosis of Bearings through Vibration Signal Using Bayes Classifiers. *Int. J. Comput. Aided Eng. Technol.* **2014**, *6*, 14–28.
28. Krishna, H. Fault Diagnosis of welded joint through vibration signals using NB Algorithm. In Proceedings of the International Conference on Advances in Manufacturing and Material Engineering, Mangalore, India, 27–29 March 2014.
29. Sugumaran, V.; Muralidharan, V.; Ramachandran, K.I. Feature selection using decision tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing. *Mech. Syst. Signal Process.* **2007**, *21*, 930–942.
30. Quinlan, J.R. Improved Use of Continuous Attributes in C4.5. *J. Artif. Intell. Res.* **1996**, *4*, 77–90.
31. Breerton, R.G.; Lloyd, G.R. Support vector machines for classification and regression. *Analyst* **2010**, *135*, 230–267.

32. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000; pp. 1–28.
33. Zhang, X.; Liang, Y.; Zhou, J.; Zang, Y. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement* **2015**, *69*, 164–179.
34. Loparo, K. *Bearings Vibration Data Set*; Case Western Reserve University: Cleveland, OH, USA. Available online: <http://www.eecs.case.edu/laboratory/bearing/welcome-overview.htm> (accessed on 20 July 2012).
35. Wu, S.D.; Wu, C.W.; Wu, T.Y.; Wang, C.C. Multi-Scale Analysis Based Ball Bearing Defect Diagnostics Using Mahalanobis Distance and Support Vector Machine. *Entropy* **2013**, *15*, 416–433.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).