

# Fault diagnosis of rolling element bearing using Naïve Bayes classifier

Xiao Jian Yi<sup>1</sup>, Yue Feng Chen<sup>2</sup>, Peng Hou<sup>3</sup>

<sup>1</sup>Department of Overall Technology, China North Vehicle Research Institute, Beijing, China

<sup>2</sup>The Fourth Research Laboratory, Beijing Special Vehicle Institute, Beijing, China

<sup>3</sup>School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China

<sup>3</sup>Corresponding author

E-mail: <sup>1</sup>yixiaojianbit@sina.cn, <sup>2</sup>yuefengch@163.com, <sup>3</sup>apolloar@163.com

Received 11 September 2017; accepted 18 September 2017

DOI <https://doi.org/10.21595/vp.2017.19153>



**Abstract.** The development of machine learning brings a new way for diagnosing the fault of rolling element bearings. However, the method in machine learning with high accuracy often has the poor ability of generalization due to the overuse of feature engineering. To address this challenge, Naïve Bayes classifier is applied in this paper. As the one of the cluster of Bayes classifiers, its ability of classification is very outstanding. In this paper, the method is provided with a detailed description for why and how to diagnose the fault of bearing. Finally, an evaluation of the performance of Naïve Bayes classifier is presented with real world data. The evaluation indicates that Naïve Bayes classifier can achieve a high level of accuracy without any feature engineering.

**Keywords:** Naïve Bayes classifier, machine learning, fault diagnosis, rolling element bearing.

## 1. Introduction

Signal based methods are proved to be effective in the study on the fault diagnosis of bearings. In signal based methods, fault feature extraction, the key in the process of the fault diagnosis, determines the quality of the diagnosis. Although the Fourier analysis can efficiently separate the vibration signals in frequency domain for fault feature extraction, the lack of information on fault feature in time domain through the Fourier analysis is inevitable. Moreover, the fault features extracted by the Fourier analysis are helpless for non-stationary signal. Fortunately, the fault features extracted include information on both time domain and frequency domain with the emergence of wavelet analysis and Empirical Mode Decomposition, which is very helpful for the diagnosis of non-stationary signal.

Compared with signal based methods, the fault diagnosis methods, adopting machine learning, are more competitive. Lots of researchers and research results support the application of machine learning in fault diagnosis with a solid theoretical foundation. Besides that, the features extracted by machine learning from data are more objective than signal based methods. Moreover, the criterion based on the accuracy of fault diagnosis is more helpful for researchers and engineers in the choice of fault diagnosis methods.

With the development of the study on machine learning, knowledge based method has attracted increasing attention again. Many methods in machine learning are used in the fault diagnosis of bearing. For example, Hidden Markov Model [1], Artificial Neural Network [2] and support vector machine [3]. Even more, some methods in deep learning are also adopted to diagnose the fault of bearings [4]. It is important that all the methods above appear superior in performance [5].

It is worth noting that the features extracted by machine learning from data are more objective than signal based methods [6]. In fact, faults feature extraction is an important step in the process of fault diagnosis. Signal separation need to be executed in faults feature extraction. Due to the signal includes the information about the system monitored, the knowledge of the system can be acquired through the signal analysis. Signal separation is an important method in signal analysis, and it maps the signal into a group of significant basis, which show the components of the signal effectively. However, the definition of the significant basis is not clear, which generates different

results using signal separation [7]. Compared with signal separation, the methods in machine learning extract the fault feature automatically in a unified standard. Therefore, the feature extracted by machine learning is objective.

The accuracy of fault diagnosis is a desirable characteristic for diagnosis systems [8]. All the desirable characteristics are standards or criteria for practitioners to evaluate and choose the different diagnosis methods. However, the accuracy of fault diagnosis is rarely mentioned in traditional method such as the Fourier analysis and wavelet analysis. In the study of machine learning, accuracy is the primary indicator of the performance of methods. Therefore, the accuracy of fault diagnosis can provide a nice guideline for the choose of diagnosis methods.

The fault diagnosis systems of bearings often require a much higher accuracy rate due to the importance of the bearings. Bearings are the frequently used units, and its failure often brings about the fatal breakdown of system [9]. More importantly, the failure caused by the fault bearings may bring about the loss of production and human casualties.

In order to improve the accuracy of diagnosis system, researchers try many methods in machine learning. However, a dilemma is exposed to practitioners. The method with high accuracy often has the poor ability of generalization [10]. The reason for the dilemma is that the overuse of feature engineering restricts the ability of generalization, although the accuracy is improved by the feature engineering. Therefore, it is necessary to find a method in machine learning to achieve the high accuracy without any or less feature engineering [11].

To address this challenge, the Naïve Bayes Classifier is proposed. The performance of the Naïve Bayes Classifier is evaluated by the vibration data of bearing that collected in real-world. The result of the evaluation demonstrates that the diagnosis performance of Naïve Bayes Classifier without feature engineering outperforms many classification technologies using feature engineering.

## 2. The principle of fault diagnosis using Naïve Bayes classifier

The theoretical foundation of Naïve Bayes classifier is Bayesian decision theory. It considers the fault diagnosis as a sequence classification in machine learning. Based on that, vibration signal is classified into different categories in sequence classification, then the fault diagnosis of bearings is finished through the category of the vibration signal.

### 2.1. The theoretical foundation of Naïve Bayes classifier

The theoretical foundation of Naïve Bayes classifier is Bayesian decision theory [10]. It thinks how to label samples optimally using posterior probability and mislabel loss. In fact, posterior probability and mislabel loss are the key elements in Bayesian decision theory. Posterior probability  $P(C_i|\mathbf{x})$  defines the probability that sample  $s$  with the attributes  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  belongs to category  $C_i \in \mathbf{C} = (C_1, C_2, \dots, C_n)$ . Mislabel loss  $\lambda_{i,j}$  describes the loss produced by the mislabel of the sample with label  $C_i$  to the  $C_j$ . The conditional risk, composed of Posterior probability  $P(C_i|\mathbf{x})$  and Mislabel loss  $\lambda_{i,j}$ , is defined as  $R(C_i|\mathbf{x}) = \sum_{j=1}^N \lambda_{i,j} P(C_j|\mathbf{x})$ . The conditional risk defines the loss produced by the classification of the sample  $s$  with the attributes  $\mathbf{x}$  to category  $C_i$ .

In the training step, the objective is to minimize the sum of the conditional risk for every sample in the data set, which results in the hypothesis  $h$  determines  $\mathbf{x}$  belongs to which category. In fact, many methods in machine learning have the common objective in training phrase, then they achieve the  $h$  through learning in the data set  $\mathbf{D} = \{s_1, s_2, \dots, s_m\}$ . The optimal hypothesis  $h^*$  is the one that minimizes the sum of the conditional risk for every sample. In Bayesian decision theory, the sum of the conditional risk for every sample in the data set is called the total risk  $R(h)$ . However, there is no need to minimize the total risk  $R(h)$ . According to the Bayes decision rule, the only thing needed is to minimize the conditional risk  $R(C_i|\mathbf{x})$  through choosing the optimal

label for  $\mathbf{x}$  to achieve the objective in training phrase, which can be represented as the formula in Eq. (1). In most literatures,  $h^*$  is also called the optimal Bayes classifier:

$$h^*(\mathbf{x}) = \arg_{C \in \mathcal{C}} \min R(C|\mathbf{x}). \tag{1}$$

Naïve Bayes Classifier is the extreme one in the cluster of Bayes classifier. In general, the mislabel loss  $\lambda_{i,j}$  in Naïve Bayes classifier is defined in Eq. (2):

$$\lambda_{i,j} = \begin{cases} 0, & i = j, \\ 1, & \text{otherwise.} \end{cases} \tag{2}$$

The aim for the definition of the mislabel loss in Naïve Bayes classifier is to minimize the error of the classification. Therefore, the Naïve Bayes classifier can be represented as:

$$h^*(\mathbf{x}) = \arg_{C \in \mathcal{C}} \max P(C|\mathbf{x}). \tag{3}$$

According to Bayes theorem,  $P(C|\mathbf{x})$  can be represented as:

$$P(C|\mathbf{x}) = \frac{P(C)P(\mathbf{x}|C)}{P(\mathbf{x})}. \tag{4}$$

$P(\mathbf{x})$  in Eq. (4) is called evidence, which is independent of the label. Therefore, Eq. (3) can be formulated as follow:

$$h^*(\mathbf{x}) = \arg_{C \in \mathcal{C}} \max P(C)P(\mathbf{x}|C). \tag{5}$$

In fact, the difficulty in the training phrase is to solve the  $P(\mathbf{x}|C)$ . However, Naive Bayes classifier adopts the attribute conditional independence assumption to overcome the difficulty. Based on the assumption,  $P(\mathbf{x}|C)$  can be represented as follow:

$$P(\mathbf{x}|C) = P(x_1, x_2, \dots, x_n|C) = \prod_{i=1}^n P(x_i|C). \tag{6}$$

Then, the formula of Naive Bayes classifier in Eq. (3) can be represented as:

$$h^*(\mathbf{x}) = \arg_{C \in \mathcal{C}} \max P(C) \prod_{i=1}^n P(x_i|C). \tag{7}$$

Obviously, the assumption, attribute conditional independence assumption, makes the calculation easy to solve. However, compared with other classifier in the cluster of Bayes classifier, the simplification of  $P(\mathbf{x}|C)$  is extreme.

## 2.2. Sequence classification

Sequence classification is a typical problem in machine learning [12]. The sequence classified is often defined as an order list of event. The event may be a symbol, a real number. At present, the sequence in sequence classification is a time series. Every sequence associates a label, and every sequence have a fixed length in order to be available to a specified classifier [13]. Therefore, the fault diagnosis of bearings using vibration signal can be categorized to sequence classification through Naïve Bayes classifier.

### 3. The procedure of fault diagnosis using Naïve Bayes classifier

The procedure of fault diagnosis using Naïve Bayes classifier is the same with the procedure of classification in machine learning. There are three steps: data clean, training classifier and classification.

#### 3.1. Data clean

The main aim of data clean is to align the sequence that needed to be classified in the data set. In fact, the performance of various classifier in machine learning is affected by the quality of the data set, however, data clean is main approach to improve the quality of the data set. There are some manipulations in data clean such as value transform, type conversion and invalid data deletion [14]. All the manipulations can improve the quality of data set effectively. In sequence classification, align the sequence is the most important thing for data clean. Because any classifier is available to the specified length sequence, it is necessary for classifiers to tailor the sequence.

#### 3.2. Training the classifier

Through the step of data clean, the data set  $D$  is represented as  $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$ . Each sample in  $D$  consists of the sequence and the label. The sequence  $\mathbf{x}_i = (x_i^1, x_i^2, \dots, x_i^M)$  is a time series has the length  $M$ , which is the vibration signal. The label  $y_i$  is the system state or the failure mode, and  $y_i \in \{C_1, C_2, \dots, C_k\}$ .

The first step in training phrase is to calculate the  $P(C_k)$ . The Eq. (8) is used to calculate it:

$$P(y = C_k) = \frac{\sum_{i=1}^N \mathbf{I}(y_i = C_k)}{N}. \quad (8)$$

The  $\mathbf{I}(\cdot)$  is the indicator function. When  $\mathbf{I}(True)$ , it equals to 1, otherwise, it equals to 0.

In the second step of training phrase, suppose that the  $k$ th element in the attribute vector  $\mathbf{x}_i = (x_i^1, x_i^2, \dots, x_i^M)$  is  $x^j$ , which has  $l$  possible values denoted by  $a_{j1}, a_{j2}, \dots, a_{jl}$ . The possibility of the attribute  $x^j$  equals to the value  $a_{jl}$  is calculated by Eq. (9):

$$P(x^j = a_{jl} | y = C_k) = \frac{\sum_{i=1}^N \mathbf{I}(x_i^j = a_{jl}, y_i = C_k)}{\sum_{i=1}^N \mathbf{I}(y_i = C_k)}. \quad (9)$$

Up to now, the training of classifier is finished through all the two steps.

#### 3.3. Classification

After the training, the Naïve Bayes classifier can be used to classify. The procedure is described below.

Step 1: calculate the possibility that the sequence belongs to category  $C_k$ :

$$P(y = C_k | \mathbf{x}) = P(y = C_k) \prod_{j=1}^M P(x^j | y = C_k). \quad (10)$$

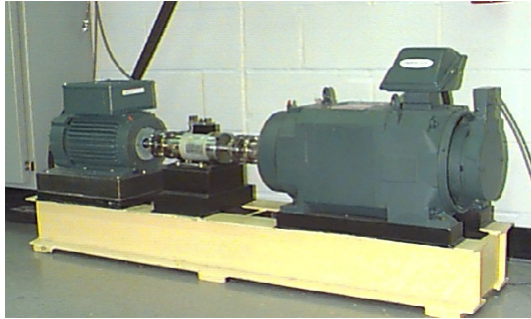
Step 2: identify the label  $y$  of the sequence:

$$y = \arg_{C_k \in \mathcal{Y}} \max P(y = C_k | \mathbf{x}). \quad (11)$$

Finally, the result of diagnosis can be determined by the  $y$ .

#### 4. Experimental verification and result

In order to test the performance of Naïve Bayes classifier on the fault diagnosis of bearings, the vibration data from Case Western Reserve Lab (CWRU) [15] is chosen. This data set has been analyzed by a number of other researchers [16-18], and those analysis results can be considered as a benchmark.



**Fig. 1.** The vibration separator

Vibration data was collected through accelerometers, and the accelerometers are placed at the 12 o'clock position at the drive end of the motor housing. Vibration signals are gathered through a 16 channel DAT recorder with sample rate 12 K/s and 48 K/s. the loads of Driver and bearing include four different types (0, 1, 2 and 3 hp), and there are four different fault types (Normal condition, ball fault, inner race fault and outer race fault). Every fault type has four fault conditions. The fault conditions are different in the fault diameters: 0.007 in, 0.014 in, 0.021 in and 0.028 in. Twenty data sets are used in this paper, which include two fault types: Normal condition and the inner race fault. All the data in this paper are sampled in rate 12 K/s. The data in Normal condition is made up of 480000 points, and all the other data have at most 125886 points.

All the twenty data sets are separated to form the new data set in a fixed length of 4096 in data clean, in order to align the sequence. Each sample in the new data set includes 4096 points, and the amounts of samples are 878. The label of each sample is equal to the data from which the sample separated. The label of the samples  $y = \{0, 1, 2, 3, 4\}$ , and the definition of the element is shown in Table 1. All of the experiments in this paper are done on intel Core i7-6700 2.60 Hz with 32GB RAM.

**Table 1.** The fault type for labels

Serial number	Label	The fault types
1	0	Normal condition
2	1	The fault diameter is 0.007 in
3	2	The fault diameter is 0.014 in
4	3	The fault diameter is 0.021 in
5	4	The fault diameter is 0.028 in

The classification result is obtained by Naïve Bayes classifier in Table 2, which is followed the process of the fault diagnosis using Naïve Bayes classifier in Section 3. The accuracy is compared with three results that achieved through SVM and SVM with feature engineering.

**Table 2.** The accuracy of different methods

Algorithm	Naïve Bayes	SVM	SVM + fractal dimension	SVM + wavelet packet
Accuracy	96 %	72 %	80.6 % [11]	94 % [19]

## 5. Conclusions

In this paper, Naïve Bayes classifier is employed to diagnose the different fault conditions of bearings. From the experimental verification and the result of fault diagnosis application in Section 4, it can be draw that Naïve Bayes classifier is effective in diagnosing the fault of bearings through the vibration signals. Furthermore, the Naïve Bayes classifier achieves a better performance than other methods with feature engineering. The proposed method not only provides a desirable characteristic: accuracy, but also provides a strong ability of generalization. Therefore, it is helpful for practitioners on trade-off between accuracy and generalization.

## References

- [1] **Liu T., Chen J., Dong G.** Singular spectrum analysis and continuous hidden Markov model for rolling element bearing fault diagnosis. *Journal of Vibration and Control*, Vol. 21, 8, p. 1506-1521.
- [2] **Weihua L. I.** A firefly neural network and its application in bearing fault diagnosis. *Journal of Mechanical Engineering*, Vol. 51, Issue 7, 2015, <https://doi.org/10.3901/JME.2015.07.099>.
- [3] **Wu S. D., Wu C. W., Wu P. H., et al.** Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Third International Conference on Mechanic Automation and Control Engineering*, IEEE Computer Society, 2012, p. 2650-2654.
- [4] **Guo X., Chen L., Shen C.** Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement*, Vol. 93, 2016, p. 490-502.
- [5] **Mao W., He J., Li Y., et al.** Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, Vol. 231, 8, p. 1560-1578.
- [6] **Lin J., Qu Liangsheng** Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of Sound and Vibration*, Vol. 234, Issue 1, 2000, p. 135-148.
- [7] **Mallat S.** *A Wavelet Tour of Signal Processing*. Elsevier, 1999, p. 83-85.
- [8] **Venkatasubramanian V., Rengaswamy R., Yin K., et al.** A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers and Chemical Engineering*, Vol. 27, Issue 3, 2003, p. 293-311.
- [9] **Muro H., Tsushima N., Nunome K.** Failure analysis of rolling bearings by X-ray measurement of residual stress. *Wear*, Vol. 25, Issue 3, 1973, p. 345-356.
- [10] **Robert C.** *Machine learning, a probabilistic perspective*. Mathematics Education Library, Vol. 58, Issue 8, 2012, p. 27-71.
- [11] **Yang J., Zhang Y., Zhu Y.** Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mechanical Systems and Signal Processing*, Vol. 21, Issue 5, 2007, p. 2012-2024.
- [12] **Xing Z., Pei J., Keogh E.** A brief survey on sequence classification. *ACM SIGKDD Explorations Newsletter*, Vol. 12, Issue 1, 2010, p. 40-48.
- [13] **Zhou P. Y., Chan K. C. C.** A feature extraction method for multivariate time series classification using temporal patterns. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer International Publishing, 2015, p. 409-421.
- [14] **Bishop C. M.** *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, New York, 2006.
- [15] **Loparo K. A.** *Bearings Vibration Data Set*. Case Western Reserve University, <http://www.eecs.cwru.edu/laboratory/bearing/download.html>.
- [16] **Lei Y., He Z., Zi Y.** A new approach to intelligent fault diagnosis of rotating machinery. *Expert Systems with Applications*, Vol. 35, Issue 4, 2008, p. 1593-1600.
- [17] **Hu Q., He Z., Zhang Z., et al.** Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. *Mechanical Systems and Signal Processing*, Vol. 21, Issue 2, 2007, p. 688-705.
- [18] **Huang Y., Liu C., Zha X. F., et al.** Research Article: A lean model for performance assessment of machinery using second generation wavelet packet transform and Fisher criterion. *Expert Systems with Applications*, Vol. 37, Issue 5, 2010, p. 3815-3822.
- [19] **Wan S., Tong H., Dong B.** Bearing fault diagnosis using wavelet packet transform and least square support vector machines. *Journal of Vibration Measurement and Diagnosis*, Vol. 30, Issue 2, 2010, p. 149-152.