Real-time Online Fault Diagnosis of Rolling Bearings Based on KNN Algorithm

Haiya Wang¹, Zhongqing Yu² and Lu Guo³

¹School of Data Science and Software Engineering, Qingdao University, Qingdao City, Shandong Province, 266071, China
²School of Data Science and Software Engineering, Qingdao University, Qingdao City, Shandong Province, 266071, China
³School of Data Science and Software Engineering, Qingdao University, Qingdao City, Shandong Province, 266071, China

*Corresponding author’s e-mail: 2017021455@qdu.edu.cn

Abstract. In order to realize the predictive maintenance of rolling bearings in industry, this paper proposes a real-time online fault diagnosis method for rolling bearings based on KNN algorithm. The method mainly includes two steps: fault diagnosis model training and real-time online fault diagnosis. Firstly, the vibration signal is preprocessed: data classification, data cleaning, data segmentation and feature parameter extraction, and then training and optimizing the fault diagnosis model. In the real-time online fault diagnosis part of the rolling bearing, the real-time online extraction of the characteristic parameters of the vibration signal is used to realize real-time online fault diagnosis through the fault diagnosis model. The results show that the fault diagnosis model based on KNN algorithm is better than the fault diagnosis model based on C4.5 algorithm and CART algorithm, which is more suitable for fault diagnosis of rolling bearings. Using this method to diagnose rolling bearings can help predictive maintenance before rolling bearing failures and reduce the economic losses caused by unplanned downtime of critical equipment.

1. Introduction
Rolling bearings are common parts of rotating machinery. Failure of rolling bearings often affects the normal operation of the equipment, resulting in unplanned downtime of the equipment and economic losses. Before the rolling bearing fails, fault diagnosis can effectively reduce the economic loss caused by unplanned shutdown. Therefore, before the failure of the rolling bearing, real-time online fault diagnosis of the bearing has a strong practical significance [1].

The fault diagnosis of rolling bearings has attracted the attention of many researchers at home and abroad [2]. Borghesani p et al. proposed a method for diagnosing bearing faults at different speeds and loads [3]. Liu Ruochen et al. proposed a method based on the extraction of electrostatic RMS as a characteristic value for bearing fault diagnosis [4]. Adil et al. proposed an exponential Discriminant Analysis (EDA) index discriminant analysis method for fault diagnosis of small samples [5]. Li Heng et al. proposed a method based on short-time Fourier transform and convolutional neural network for fault diagnosis of bearings [6]. Zhang Qiantu et al. proposed a bearing fault diagnosis method based on SVM [7]. Zhou Yong et al. proposed a bearing fault diagnosis method based on RVM [8]. Zhang et al. proposed a bearing fault diagnosis method based on deep convolutional neural network. The method is
still good in fault diagnosis under variable load and noise environment [9]. Based on the amplitude of the bearing vibration signal, Janssens et al. used CNN to realize the fault diagnosis of rolling bearings, and the diagnostic effect is good [10]. The above method has certain limitations in the application of fault diagnosis for rolling bearings in industry. For CNN neural networks, parameters such as the number of convolutional layers and convolution kernels of CNN for specific problems are difficult to determine [9]; SVM and RVM are sensitive to parameter adjustment and kernel function selection, and the training time complexity is high. How to apply the fault diagnosis algorithm to the industry to achieve fault diagnosis? Aiming at the above problems, this paper proposes a real-time online fault diagnosis method for rolling bearings based on KNN algorithm.

2. Bearing Fault Diagnosis Model Construction

The fault diagnosis model of the bearing is composed of two parts: the first part is to preprocess the historical data of the bearing vibration signal, make these data suitable for the establishment of the rolling bearing fault diagnosis model; the second part is based on the pre-processed data and algorithm, training optimization Fault diagnosis model.

2.1 Data Preprocessing

The vibration signal data processing of the rolling bearing includes: data classification, data cleaning, data segmentation, feature extraction of data, and data filtering and smoothing. In order to effectively eliminate and suppress impurities in the data set, a sliding mean filtering algorithm is adopted, and its core formula is: \[ \bar{y}_n = \frac{1}{N} \sum_{i=0}^{N-1} x_{n-i}, \] \[ y_n \] is the output of the nth sampling [11].

2.2 Model Generation

Fault diagnosis model generation of rolling bearings includes model training and model optimization based on historical data.

2.2.1 Model Generation Based on KNN Algorithm

The KNN algorithm is a basic classification and regression method.

The algorithm is described as follows: Input: training data set \( T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}, x_i \in \mathbb{R}^p, y_i \in \{c_1, c_2, \ldots, c_k\} \) And test data x output: the category to which instance x belongs.

1. According to the given distance metric, find the k samples closest to the x distance in the training set \( T \), and the neighborhood of \( x \) covering the k points is recorded as \( N_k(x) \). Commonly used is the Minkowski distance, defined as: \[ D(x, y) = \left( \sum_{i=1}^{m} |x_i - y_i|^p \right)^{\frac{1}{p}} \] (2)\[ \text{Where } P \geq 1 \text{ is the Euclidean distance when } p = 2, \text{ and Manhattan distance when } p = 1. \]

2. Determine the category \( y \) of \( x \) in \( N_k(x) \) according to the classification rules (such as majority vote):

\[ y = \arg \max_j \sum_{x \in N_k(x)} I[y_i = c_j], \quad i=1,2, \ldots, n_j=1,2,\ldots,K \] (3)

Given a test sample \( x \), its nearest \( k \) training instances form a set \( N_k(x) \), and if the category covering the \( N_k(x) \) region is \( c_j \), the classification error rate: \[ \frac{1}{k} \sum_{x \in N_k(x)} I[y_i \neq c_j] = 1 - \frac{1}{k} \sum_{x \in N_k(x)} I[y_i = c_j] \]

The choice of \( k \), the choice of \( k \) will have a significant impact on the results of the algorithm. A cross-validation method is usually used to select the appropriate \( k \) value [12-13]. According to the KNN algorithm principle and bearing historical data, the bearing fault diagnosis model based on KNN algorithm is generated and optimized.

2.2.2 Model Generation Based on C4.5 Algorithm

The C4.5 algorithm uses the post-prune technique to pruning the decision tree generated based on historical data, generates a decision tree model, and generates a series of IF-THEN rules according to the model to realize the classification of the data set. The core formula of the C4.5 algorithm is: [14-15].

1. Information entropy, let \( S \) be a collection of \( S_i \) data samples, Assuming that the class label
attribute has \( m \) different values, \( m \) different classes \( T_i \) (\( i = 1, ..., m \)) are defined. Let \( S_i \) be the number of samples in class \( T_i \). The expected value for a given sample classification is:

\[
\text{Info}(S) = -\sum_{i=1}^{m} \frac{s_i}{S} \log_2 \frac{s_i}{S} \tag{5}
\]

2. Information gain, the amount of information divided into subsets by attribute \( A \) is:

\[
E(A) = \text{Info}(S) - \sum_{j=1}^{n} \frac{S_i}{S} \cdot \text{Info}(S_{ij}) \tag{6}
\]

3. Information gain rate, information gain rate is introduced in the C4.5 algorithm.

\[
\text{GainRatio}(A) = \frac{\text{Gain}(A)}{\text{SplitE}(A)} \tag{7}
\]

\[
\text{SplitE}(A) = \sum_{i=1}^{k} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \tag{8}
\]

According to the C4.5 algorithm principle and bearing historical data, the bearing fault diagnosis model based on C4.5 algorithm is generated and optimized.

2.2.3 Model Generation Based on CART Algorithm

The CART algorithm is similar to the ID3 algorithm, but the CART algorithm can only establish a binary tree. The CART algorithm uses the Gini index \( G_i \) to measure purity. The attribute is obtained as the root node based on the Gini index, and then the model is constructed by the top-down method in a recursive manner until each sample set is completely pure after the division, and the tree construction is stopped. For sample set \( D \), the definition is as shown in equation (9) [16].

\[
G_i(D) = 1 - \sum_{i=1}^{N} p_i^2 \tag{9}
\]

Where \( N \) is the number of class labels and \( p_i \) is the probability that category \( i \) appears in the sample set \( D \). According to the CART algorithm principle and bearing historical data, the bearing fault diagnosis model based on CART algorithm is generated and optimized.

3. Bearing Real-time Online Fault Diagnosis Process

Using the historical data of the pre-processed rolling bearing, the fault diagnosis model of the rolling bearing is generated and optimized, which can effectively realize the real-time online fault diagnosis of the rolling bearing.

4. Experimental Verification

In order to verify the feasibility of the real-time online fault diagnosis model method based on KNN algorithm, two sets of data sets and two algorithms were used for comparison verification. The experimental data was obtained from the data center website of the Western Reserve University. Four kinds of status data of the bearing drive end were selected, which were normal state, inner ring fault, ball fault, and outer ring fault.

Data set 1: Select 12k (diameter 0.007; load 1; motor speed 1772) drive end bearing fault data, the data set has 849,600 pieces of data. The data was preprocessed, with 100 raw data as one sample, for a total of 8496 samples. The sample sizes of the rolling bearing normal state, inner ring fault, ball fault, and outer ring fault were 4839, 1219, 1214, and 1224, respectively.
4.1 Vibration Signal
The time domain diagram and frequency domain diagram based on the original vibration data of the rolling bearing include the normal state of the rolling bearing drive end, the inner ring fault, the ball fault and the outer ring fault signal.

![Figure 2. The normal state signal](image)

![Figure 3. The inner ring fault signal](image)

![Figure 4. The ball fault signal](image)

![Figure 5. The outer ring fault signal](image)

4.2 Data Preprocessing
The preprocessing of the vibration signal data of the rolling bearing includes: data classification, data cleaning, data segmentation, feature extraction of the data, and data filtering and smoothing, which finally meets the needs of generating and optimizing the fault diagnosis model. Figure 6 is part of the data after pre-processing of the original data, and Figure 7 is part of the data after feature extraction of the original data:

![Figure 6. Partial data after raw data preprocessing](image)

![Figure 7. Partial data after feature extraction](image)

4.3 Real-time Online Fault Diagnosis of Rolling Bearings
The model training data of the three algorithms accounted for 75% of the total data, and the test data accounted for 25% of the total data. Figure 8 is a statistical diagram of the fault diagnosis accuracy rate based on the C4.5 and CART algorithms. Figure 9 is a rendering of the real-time online fault diagnosis model based on the KNN algorithm:
According to the time domain diagram, frequency domain diagram and historical data of the rolling bearing, the real-time online fault diagnosis results of the bearing based on KNN algorithm are compared. It is verified that the fault diagnosis model can accurately diagnose the bearing in real time based on the bearing state data. To further verify the practicality of the bearing fault diagnosis method, another set of rolling bearing data sets were used for re-verification.

Data set 2: Select 1.1k (diameter 0.007; load 1; motor speed 1772) drive end bearing fault data, select 80,000 data sets, preprocess the data, take 100 data as a sample, a total of 800 samples. The sample size of the rolling bearing normal state, inner ring fault, ball fault, and outer ring fault is 200.

The model training data of the three algorithms accounted for 75% of the total data, and the test data accounted for 25% of the total data. Figure 10 is a statistical diagram of the fault diagnosis accuracy rate based on the C4.5 and CART algorithms. Figure 11 is a rendering of the real-time online fault diagnosis model based on the KNN algorithm.

According to the time domain diagram, frequency domain diagram and historical data of the rolling bearing, the real-time online fault diagnosis results of bearing based on KNN algorithm are compared, and the feasibility of the fault diagnosis method is further verified.

4.4 Comparative Analysis

Table 1 and Table 2 respectively count the diagnosis results of two data sets and three fault diagnosis models.

| Bearing status | Normal state | Inner ring failure | Ball failure | Outer ring failure |
|----------------|--------------|--------------------|--------------|--------------------|
| Number of samples | 4839         | 1219               | 1214         | 1224               |
| C4.5 algorithm    | 94.5160%     | (Average accuracy of 100 tests) |
| CART algorithm    | 90.8291%     | (Average accuracy of 100 tests) |
| KNN algorithm     | 96.1%        | (Average accuracy of 100 tests) |
Table 2 statistics based on the fault model diagnosis results of data set 2

| Bearing status          | Normal state | Inner failure | Ball failure | Outer failure |
|-------------------------|--------------|---------------|-------------|--------------|
| Number of samples       | 200          | 200           | 200         | 200          |
| C4.5 algorithm          | 94.4129%     | (Average accuracy of 100 tests) |
| CART algorithm          | 90.7264%     | (Average accuracy of 100 tests) |
| KNN algorithm           | 95.3%        | (Average accuracy of 100 tests) |

By comparing Tables 1 and 2, the analysis can be found:
1. The fault diagnosis accuracy rate based on KNN algorithm model in Table 1 and Table 2 is higher than that based on C4.5 algorithm and CART algorithm.
2. Rich and complete historical data is conducive to the optimization of the rolling bearing fault diagnosis model.

5. Concluding Remarks
The results show that the real-time online fault diagnosis method based on KNN algorithm can accurately realize the real-time online diagnosis of rolling bearing faults, and the fault diagnosis accuracy based on KNN algorithm is better than C4.5 algorithm and CART algorithm. The real-time online fault diagnosis method of rolling bearing based on KNN algorithm is more suitable for fault diagnosis of rolling bearings; it has strong practical application significance for real-time online fault diagnosis of rolling bearings in industry.

REFERENCES
[1] Yiwen Wang, Study on the Fault Diagnosis and Prediction System of Rolling Bearing Based on Vibration Signal [D]. China University of Mining and Technology, 2014.
[2] Yan Wang, et al. Bearing Fault Diagnosis Method Based on Variational Bayes. Computer Science. https://kns.cnki.net/KCMS/detail/50.1075.TP.20190822.0849.002.html
[3] BORGHESANI, et al. A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions[J]. Mechanical Systems & Signal Processing, 2013, 38(1): 23-35.
[4] LIU Ruochen, ZUO Hongfu. Research on electrostatic monitoring method of rolling bearings with injected fault under variable operating conditions[J]. Chinese Journal of Scientific Instrument, 2014, 35(10): 2348-2355.
[5] ADIL M, ABID M, KHAN A Q, et al. Exponential discriminant analysis for fault diagnosis[J]. Neurocomputing, 2016, 171: 1344-1353.
[6] LI Heng, ZHANG Qin, et al. Fault diagnosis method for rolling bearings based on short-time fourier transform and convolution neutral network[J]. Journal of Vibration and Shock, 2018, 37(19): 124-131 (In Chinese).
[7] ZHANG QianTu, et al. Mechanical Fault Diagnosis Based on LCD Information Entropy Feature and SVM[J]. Journal of Mechanical Transmission, 2015, 39(12): 144-148.
[8] ZHOU Yong, HE ChuangXin. Bearing fault diagnosis under varying load conditions based on individual feature selection and relevance vector machine[J]. Journal of Shock and Vibration, 2012, 31(3): 157-161.
[9] Zhang W, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load [J]. Mechanical Systems & Signal Processing, 2017, 100: 439-453.
[10] Olivier Janssens, et al. Convolutional neural network based fault detection for rotating machinery[J]. Journal of Sound and Vibration, 2016(337): 331-345.
[11] ZHOU Jiyu et al. Design of electronic scale based on mean sliding filtering algorithm and STM32 [J]. Modern Electronics Technique, 2017. 40(10): 10-15.
[12] Bingnan Liu, et al. High resistance grounding fault identification in distribution network based on
[13] Zhinong Jiang, et. al. Operating Mode Identification Method for Diesel Engines based on Manifold Learning and KNN Algorithm [J]. Noise and Vibration Control, 2019, 39(03): 1-6.

[14] Cunge Han, et al. Research and Improvement of C4.5 Algorithm in Decision Tree Classification Algorithm [J]. Computer Systems & Applications, 2019, 28(06): 198-202.

[15] Yinjuan Dai, et al. Research on New Vehicle Fault Prediction Method Based on Improved C4.5 Algorithm [J]. Journal of Changshu Institute of Technology, 2019, 33(05): 72-77.

[16] Zhinong Jiang et al. Research on Diesel Engine Fault Diagnosis Method Based on CART Decision Tree [J]. Journal of Beijing University of Chemical Technology (Natural Science Edition), 2018, 45(04): 71-75.