

Bearing Fault Recognition Based on Feature Extraction and Clustering Analysis

Xin Zhang^{1, a}, Jianmin Zhao^{1, b}, Haiping Li^{1, c} and Fucheng Sun^{1, d}

¹ Mechanical Engineering College, Shijiazhuang, 050003, China.

^azxyx361@163.com, ^bjm_zhao@hotmail.com, ^chp_li0929@163.com, ^dsfcghdl@163.com

Keywords: Clustering analysis, bearing, fault pattern, time domain feature parameters.

Abstract. In this paper, the clustering analysis is used to distinguish bearing fault pattern. Some time domain feature parameters are extracted from vibration signal, and the combination of three feature parameters are chosen from these feature parameters for the clustering analysis. The Euclidean distance is used to calculate the distance of point-to-center. After validation, the effect of clustering analysis is effective to distinguish the bearing fault pattern, and the best combination of feature parameters for fault pattern recognition by clustering analysis is found.

1. Introduction

Rolling bearing is an important part of industrial which is the most widely used industrial machine elements. And bearing is the most easily damaged mechanical parts. In the rotating machinery using the rolling bearing, about thirty percent of the mechanical failures are caused by the bearing. So it is important to evaluate the running state of the bearings.

A lot of scholars have carried on a more thorough research to the fault diagnosis of bearing. Signal processing method based on time-frequency domain analysis is applied to the bearing fault diagnosis. But these methods are not always effective when the data is complex. So the artificial intelligence methods are used in the bearing fault diagnosis, such as artificial neural network, support vector machine, Bayesian network and so on. These years, the clustering analysis is also used in the field of fault diagnosis, and clustering analysis is an important part of data mining. There are some scholars make research on the clustering analysis for the mechanical fault diagnosis [1-4].

The rest of the paper is organized as follows. An overview of the proposed method in this paper is performed in Section 2. Section 3 presents a illustrative case to test the effectiveness of the method. Summary follows in Section 4.

2. Proposed Method

In this section, the flow chart of the proposed method is shown as Figure 1. The first step is feature extraction and combination. The second step is clustering analysis. The third step is calculation of the recognition rate. And the last step is chose the best combination.

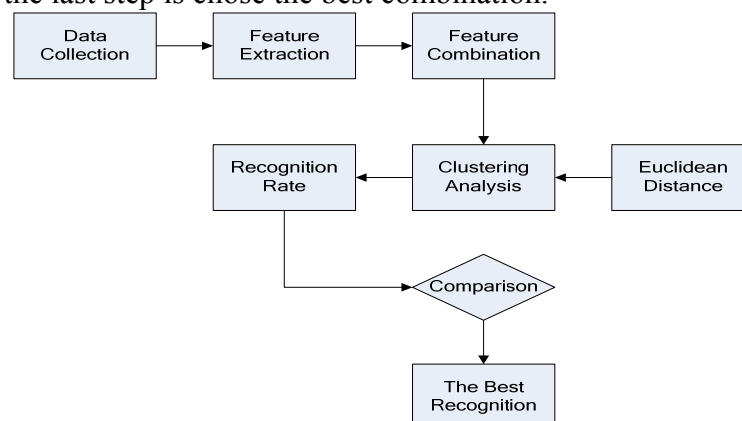


Fig. 1 The flow chart of paper

2.1 Feature Extraction

In this paper, there are some time domain feature parameters which represent the characteristic of failure are extracted for clustering analysis. These parameters calculation formulas are shown as following.

The peak to peak value x_{p-p} refers to the subtraction between the maximum amplitude value and negative amplitude maximum value in discrete vibration.

$$x_{p-p} = x_{\max} - x_{\min} \quad (1)$$

The kurtosis x_{kur} is the statistics value to explain distribution character of the vibration signal, and it is formulated as

$$x_{kur} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\left(\frac{1}{N} \sum_{i=1}^N x_i^2\right)^2} \quad (2)$$

The impulsion index I is defined as the ratio of signal peak value and mean amplitude, and it is formulated as

$$I = \frac{\max |x_i|}{\frac{1}{N} \sum_{i=1}^N |x_i|} \quad (3)$$

The margin index L is defined as the ratio of signal peak value and root amplitude, and it is formulated as

$$L = \frac{\max |x_i|}{\left(\frac{1}{N} \sum_{i=1}^N |x_i|^{1/2}\right)^2} \quad (4)$$

The clearance factor Cf is defined as the ratio of signal maximum value and the mean square value, and it is formulated as

$$Cf = \frac{x_{\max}}{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (5)$$

In the above formulas, x_i represents discrete vibration signal. And three feature parameters are used to clustering analysis which is chosen from these parameters randomly after signal feature extraction.

2.2 Basic Theory of k -means Method.

In this section, the theory of k -means for clustering analysis is explained. The k -means algorithm is the most well know and fast method in non-hierarchical cluster algorithms. And it is used in various fields for the simplicity of k -means algorithm. The k -means method is a partitioning clustering way that separates data into k mutually excessive groups. The k -means method is one of dynamic clustering algorithm. First, the k points are chosen as initial starting cluster center, and the distance from each data to each center is calculated. [5] Then the minimum distance sample is classified as nearest cluster center as shown in the Figure 2.

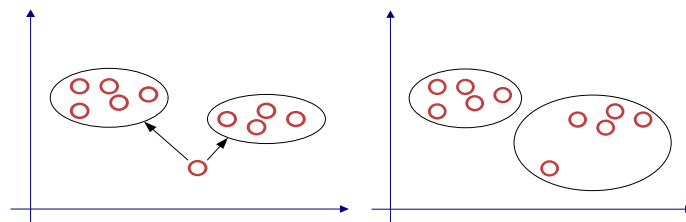


Fig. 2 Sample clustering process

Then the new sample is classified in cluster center, and the new distance center is calculate again until the new distance center equal to the last distance center. The k -means algorithm is very popular because of its ability to cluster a kind of huge data, and also outliers, quickly and efficiently. The calculate method of distance will be discuss in the next section.

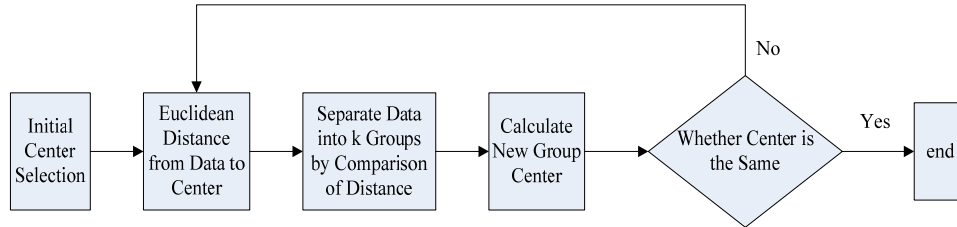


Fig. 3 The flow chart of k -means algorithm

2.3 Method of Distance Calculation.

In this section, the theory of Euclidean distance is explained. The Euclidean distance measures the straight line distance between two points. And the Euclidean distance is used to measures the difference in the individual magnitudes of each sample. The Euclidean distance is used to calculate the distance of point-to-center. The Euclidean distance is a common definition of distance and it is the real distance between two points in the m dimensional space. [6] The Euclidean distance between x_i and y_i is formulated as:

$$d = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (6)$$

In the above formula, the x_i and y_i represent two points in the m dimensional space. The m represent dimensional of space.

3. A Case Study

In this paper, the data using to validate the proposed method is the bearing data of Case Western Reserve University. As shown in Figure 4, the test stand of Case Western Reserve University consists of a 2 HP motor, a torque transducer, a dynamometer and control electronics. The test bearings support the motor shaft. Single point faults were introduced to the test bearings using electro-discharge machining with diameters of 7mils, 14mils, 21mils and 28mils. The fault location covers the inner raceway, outer raceway and rolling elements, the load variation range is 0-3 HP and the speed variation range is 1730-1797rpm. The signal sampling frequencies are 12kHz and 48kHz [7]. There are two types of bearing in this experiment, SKF 6205 and SKF 6203 bearings, and drive end bearing is SKF 6205, fan end bearing is SKF 6203.

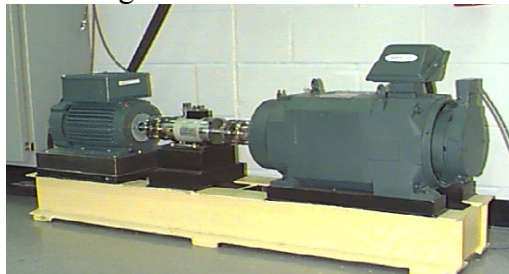
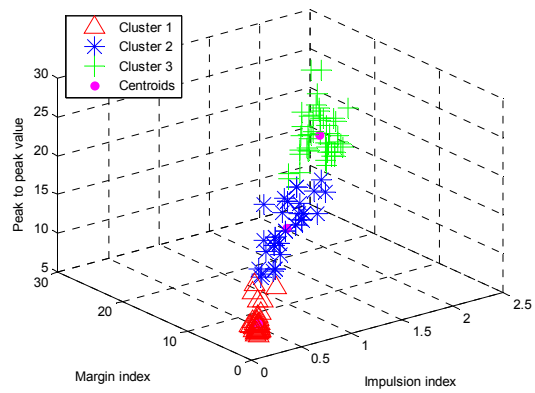
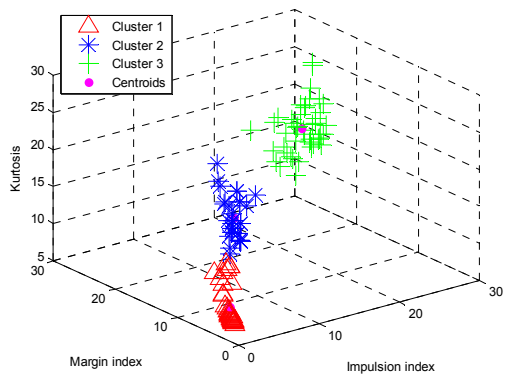
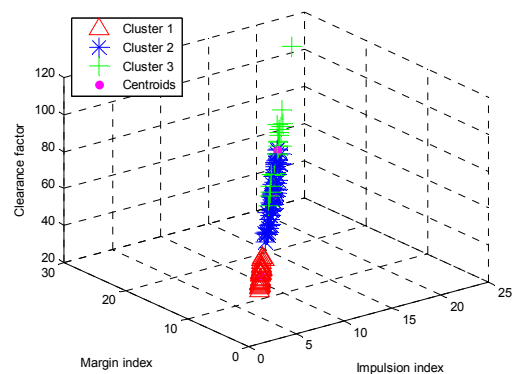
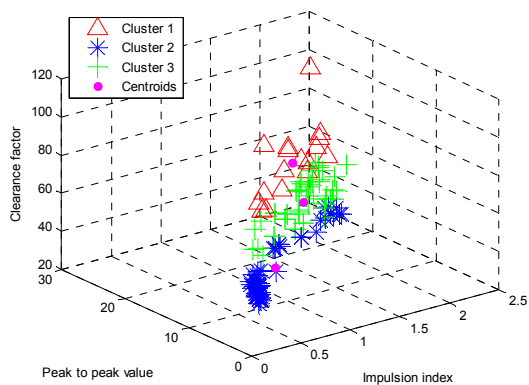
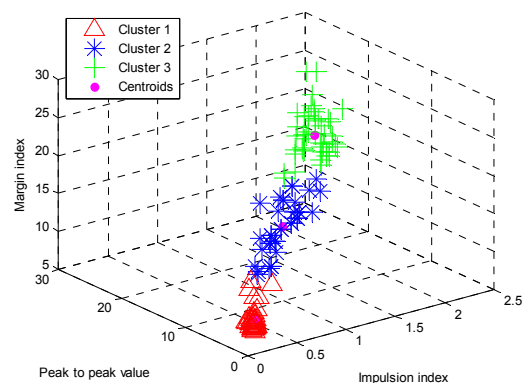
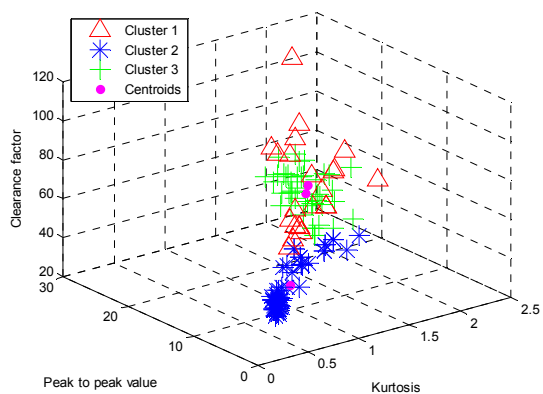
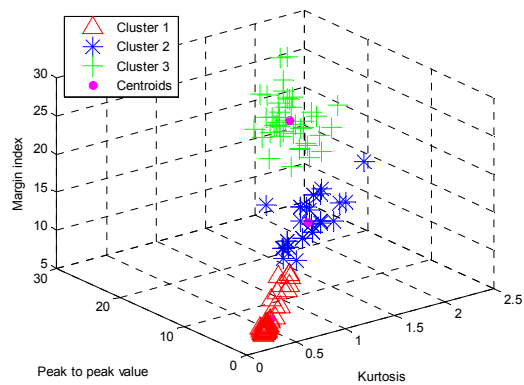
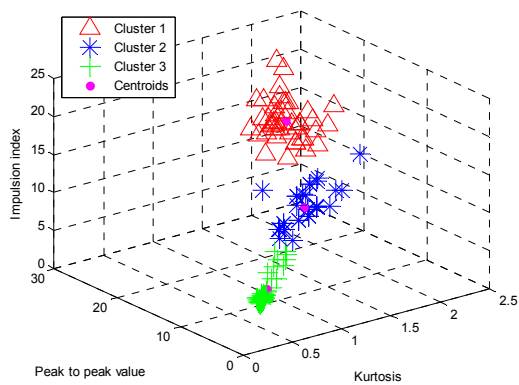


Fig. 4 Bearing test stand

In this paper, the 14mils data set is used to validate the k -means method. The feature parameters are extracted from the vibration signal. Three feature parameters are chosen as the feature sample for the clustering analysis. And compared the effect of clustering analysis to chose the best parameters group.



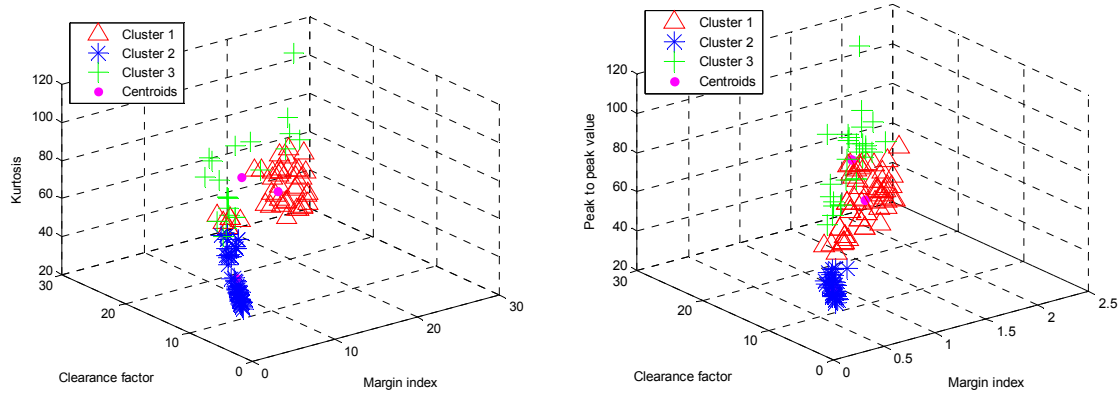


Fig. 5 Cluster analysis

And the result of the clustering analysis is shown in the Figure 5. And the fault pattern recognition rates are shown as Table 1. According the table, the best feature parameter group is kurtosis, peak to peak value and margin index because of highest recognition rate for each fault pattern.

Table 1 Three Scheme comparing

Numble	Parameter1	Parameter2	Parameter3	Ball rate	Inner rate	Outer rate
1	x_{kur}	x_{p-p}	I	80%	100%	100%
2	x_{kur}	x_{p-p}	L	90%	100%	100%
3	x_{kur}	x_{p-p}	Cf	37.5%	90%	90%
4	x_{p-p}	I	L	80%	97.5%	92.5%
5	x_{p-p}	I	Cf	27.5%	67.5%	82.5%
6	I	L	Cf	40%	72.5%	82.5%
7	I	L	x_{kur}	80%	100%	100%
8	I	L	x_{p-p}	80%	97.5%	92.5%
9	L	Cf	x_{kur}	37.5%	90%	90%
10	L	Cf	x_{p-p}	40%	80%	82.5%

4. Summary

In this paper, the feature parameters are extracted from vibration signal. And the three time domain feature parameters are chose from ten parameters as the sample point. The k -means method is used to clustering analysis for bearing fault pattern recognition. The Euclidean distance is used to calculate the distance of point-to-center. After validation, kurtosis, peak to peak value and margin index is the best group to recognize the fault pattern.

References

- [1]. Jain A. K, Data clustering: 50years beyond K-means. Pattern Recognition Letters **31**(2010), 651-666.
- [2]. Erisoglu M, Calis N, Sakallioglu S, A new algorithm for initial cluster centers in k-means algorithm. Pattern Recognition Letters **32**(2011), 1701 – 1705.
- [3]. Gao Jinxin, Hitchcock D B, James-Stein shrinkage to improve k-means cluster analysis. Computational Statistics and Data Analysis, **54**(2010) , 2113 – 2127.
- [4]. Yiakopoulos C T, Gryllias K C, Antoniadis I A. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. Expert Systems with Applications **38**(2011), 2888-2991.
- [5]. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, The people's posts and Telecommunications Press, Beijing, 2011.

- [6]. Jiawei Han, Micheline Kamber, Jian Pei, Data Mining Concepts and Techniques, China Machine Press, Beijing, 2012.
- [7]. Information on: <http://csegroups.case.edu/bearingdatacenter/home>