Health state assessment of rolling bearings based on EMD-improved Mahalanobis-Taguchi system

Xiangjin Dong*

School of Economics and Management, Nanjing University of Science and Technology Jiangsu, China

*Corresponding author: 2580726404@qq.com

Keywords: Rolling Bearing, Health State Assessment, Mahalanobis-Taguchi System, Empirical Mode Decomposition

Abstract: In order to improve the accuracy of the health state assessment of rolling bearings. A method of health state assessment based on Empirical Mode Decomposition and improved Mahalanobis-Taguchi system (IMTS) is proposed. Firstly, the modified method first uses Empirical Mode Decomposition (EMD) to extract characteristic variables from the vibration signal of the rolling bearings. Secondly, the improved Mahalanobis-Taguchi System (MTS) was used to determine and optimize the reference space, and Health Index (HI) was proposed to realize the quantitative evaluation of the health state of rolling bearings. The results show that this method can accurately and effectively evaluate the health state of rolling bearings.

1. Introduction

Prognostics health management (PHM) refers to [1] all kinds of physical parameters of key components of mechanical equipment are used in various intelligent models or mathematical algorithms to determine the operation state of the system itself. The development process of PHM technology has changed from passive measures to active prevention, and to prior maintenance and comprehensive health management. The improvement of PHM technology is based on the original technology of condition monitoring and fault diagnosis.

In recent years, scholars in academia have proposed different state assessment models combined with a variety of pattern recognition methods. In order to solve the problem that Tso methods focus on fault detection and isolation and they cannot provide effective guidance for testability design to improve PHM performance level, Liao et al [2] proposed a Tso model for PHM system. Li et al [3] improved the health index based on KPCA and EWMA and used the accumulation of hidden layers to build hierarchical gated recursive cell network. This method is used to do the actual test of bearing data and health prediction. Jiang [4] proposed a bearing performance degradation evaluation method based on HMM and nap. Zhang Gang et al [5] proposed a virtual health factor construction model based on specific frequency information entropy and SEI-RBM. Xiao Wenbin [6] proposed an evaluation model based on wavelet packet transform and hidden Markov model which can effectively predict faults. The above literature focuses on the construction of health state function. Therefore, how to extract effective feature parameters from vibration signals and optimize high-dimensional feature parameters is a problem that needs further study.

This paper presents a health state assessment method based on empirical mode decomposition (EMD) and improved Mahalanobis-Taguchi system (MTS). EMD is used to extract feature vectors, which is combined with the advantages of rough set. EMD is introduced into the Mahalanobis-Taguchi system model to screen variables instead of orthogonal array and signal-to-noise ratio. The contribution rate of different characteristic variables of Mahalanobis distance is different, so we weighted Mahalanobis distance by random forest which is an objective weighting method. Combining the above methods, the health index is proposed to evaluate the health state of rolling bearing.

The paper is organized as follows: Section 1 introduces the related research progress and content structure, Section 2 introduced some basic theories, including EMD feature extraction, MTS, IMTS and HI. Section 3 proposes assessment model carries out data experimental analysis to verify the effectiveness of this method, section 5 is the conclusion of the paper.

2. Methodology

2.1 EMD feature extraction

As an adaptive time-frequency analysis method, EMD can adaptively decompose the signal on the time and frequency domain by analyzing the local time-varying characteristics. EMD can also decompose the change trend or fluctuation of different scales in the original signal step by step. This series has different the time series of characteristic scale is decomposed into several IMF components (intrinsic mode function) [7] [8]. The specific process of EMD's decomposition of any signal is as follows:

(1) Search out all extreme points of the signal $x(t)$;

(2) Perform 3 spline interpolations on the extreme points to obtain the upper and lower envelopes, record the mean value of the upper and lower envelopes as $m(t)$, and record the original signal $x(t)$ minus $m(t)$ as $h(t)$;

(3) Compare the two conditions of the IMF to observe $h(t)$. If the conditions are met, take $h(t)$ as the first IMF component $c_1(t)$ .If at least one of the conditions is not met, then for the operations of the $h(t)$ loop steps (1) to (3), when the IMF conditions are met, the loop ends;

(4) Extract $c_1(t)$ from $x(t)$ separately and record $x(t)-c_1(t)$ as $r(t)$,Determine whether $r(t)$ can continue to extract the IMF component, if so, perform the operations of $r(t)$ loop steps (1) to (4) to obtain the second IMF component $c_2(t)$ that meets the conditions, otherwise terminate;

(5) After multiple extractions, if $r(t)$ becomes a monotonic function, the algorithm ends.

$$x(t)=\sum_{i=1}^{n} c_i(t)+r(t) \tag{1}$$

Where $c_i(t)$ represents each IMF component; n represents the number of IMF components obtained by decomposition; $r(t)$ represents the margin which represents the average trend of the original signal.

2.2 Mahalanobis-Taguchi System

The Mahalanobis-Taguchi System (MTS) is a multivariate statistical method with a deep theoretical foundation and a wide range of application scenarios, including feature reduction and classification prediction [9].This method was proposed by Genichi Taguchi, and once it was proposed, it has attracted the attention and research of many scholars at home and abroad [10].

The basic steps of MTS are:

Step 1: Determine the reference space

Let F contain $n \times p$ samples, n is the number of samples, p is the number of variables of samples, and the value of the j variable of the i sample is $x_{ij}, i=1,2,\dots,n; j=1,2,...,p$,Standardize F, the matrix element becomes $z_{ij}=(x_{ij}-\overline{x}_j)/s_j$, where $\overline{x}_j$ is the mean value of variable j, and s_j is the standard deviation of variable j. The vector of sample normalization is denoted as $Z_i=(z_{i1}, z_{i2},..., z_{ip})^T, i=1,2,...,n$.

Step 2: Verify the validity of the reference space

First, the abnormal sample data is standardized, using the mean and standard deviation of the normal sample. Then, the Mahalanobis distance of abnormal samples is calculated, and the correlation coefficient matrix of normal samples is used. Finally, if the Mahalanobis distance of abnormal samples is significantly greater than that of normal samples, the benchmark space constructed is effective, otherwise, the samples need to be re selected to construct an effective Mahalanobis space.

Step 3: Optimize the base space

Generally speaking, the larger the information gain difference Δ is, it indicates that the variable helps to improve the classification accuracy. The calculation formula of the information gain difference is:

$$\Delta = SNR^{+} - SNR^{-} \tag{2}$$

If Δ is greater than 0, it should be reserved; if Δ is less than 0, it should be discarded.

Step 4: Calculate Mahalanobis distance

The correlation coefficient matrix between variables is denoted as R, R^{-1} represents its inverse matrix, and the Mahalanobis distance calculation formula for the i sample is:

$$MD_i = \frac{1}{p} Z_i^T R^{-1} Z_i \tag{3}$$

2.3 Improved Mahalanobis-Taguchi System

(1) Benchmark space optimization based on rough set theory

The rough set method is used as a method to replace the orthogonal table and the SNR ratio to optimize the reference space [11]. The basic concept of rough set, let $U \neq 0$ be a limited set of research objects. A concept family $F = \{X_{1,} X_2, \ldots X_n\}$ in U is called knowledge about U universe. Where $X_i \in U$, $X_i \notin \phi$, $X_i \cap X_j = \phi$,

$i \neq j, i, j = 1,2, \ldots n$, $\bigcup_i X_i = U$.

Let R be an equivalence relation family, if $r \in R$ and $IND(R - \{r\}) = IND(R)$, then r can be reduced in R. If $P = R - \{r\}$ is independent, then P is a reduction in R [12].

In general, the equivalence relationship family R contains multiple attributes. Some redundant attributes r not only do not help the classification rules, but will increase the workload of information collection and reduce work efficiency.

(2)Weighted Mahalanobis distance with random forest weights

The weight of each feature in Mahalanobis distance subject is the same, that is to say, the influence of each feature on Mahalanobis distance is the same. However, in actual calculations, different characteristics will have different effects on the Mahalanobis distance, so the weighted Mahalanobis distance makes the calculation result more accurate. Chang Zhipeng et al [13] developed a kind of weighted Mahalanobis distance, which has sufficient subjective consideration but insufficient objective consideration. Su et al [14] constructed weighted Schmidt orthogonal Mahalanobis distance, weighted different features in Mahalanobis distance according to SNR gain.

The calculation formula of the weighted Mahalanobis distance is:

$$MD_i = \frac{1}{p} Z_i^T W R^{-1} W Z_i \tag{4}$$

Where: $0 \leq w_j \leq 1$ in the $W = diag[w_1, w_2, \cdots, w_p],$ weight matrix represents the weight value of the j rd variable, and satisfies $\sum_{j=1}^{p} w_j = 1$.

Random forest is a kind of sorter. Its principle is to use multiple tree classification to analyze samples. The sorter was first proposed by Leo Breman. It mainly uses resampling method to extract multiple groups of samples from a large number of initial data and construct tree classification, and selects all the classified trees to determine the final result [15].Because the objective weighting method is more suitable for the health state evaluation model [16], this paper uses the weight obtained by random forest algorithm to improve Mahalanobis distance, and constructs the health state evaluation model of rolling bearing, which provides a new idea for the weighting method of weighted Mahalanobis distance.

2.4 Health Index

After optimizing the constructed reference space, a Health Index (HI) can be constructed based on the Mahalanobis distance between the sample to be tested and the reference space, thereby quantifying the current health status of the monitored object. The value of the health index calculated on the basis of feature extraction of the vibration signal containing the running state information of the rolling bearing is often defined between 0-1, 1 represents the best running state of the rolling bearing, and 0 represents the complete failure of the bearing.

Since the value of Mahalanobis distance is between $[0,+\infty)$, this article combines the idea of distance-based evaluation with parameter analysis, uses the $\arctan x$ function, and uses the Mahalanobis distance between the sample to be tested and the optimized reference space as the independent variable. A new method of measuring health status. The formula of the health status index is as follows:

$$HI = 1 - 2 \times \arctan(\alpha \times MD_w) / \pi \tag{5}$$

Among them is the adjustment factor to ensure that the value of A is between [0, 1], then there is:

$$\alpha = \frac{\arctan(1 - HI_{ini})}{MD_{ini}} \tag{6}$$

3. Health assessment model and case analysis

3.1 Health assessment model

It is the key link of rolling bearing health management to monitor and evaluate the whole life cycle operation process of rolling bearing. Through different monitoring methods, and according to the statistical value of monitoring data to identify the health status of rolling bearing in the process of operation, complete the construction of rolling bearing health status evaluation model. Then, according to the bearing state, the corresponding maintenance or replacement decision is made to ensure its normal and stable operation. The health status assessment based on the improved Martin system and EMD is mainly divided into the following steps:

Step 1: EMD feature extraction and feature parameter construction;
Step 2: Establish and verify the validity of the reference space;
Step 3: Rough set optimizes the reference space;
Step 4: Calculation of the weighted Mahalanobis distance throughout the life cycle;
Step 5: Calculation of health assessment index;

3.2 Case analysis of EMD-IMTS health assessment model

This article uses experimental bearing life cycle data provided by the Intelligent Maintenance Center of the University of Cincinnati.

The first step of the Mahalanobis-Taguchi System is to confirm the reference space, in order to determine the amount of normal sample data and abnormal sample data of the Mahalanobis-Taguchi System, According to the change of root mean square (RMS) with time, this paper draws the development trend curve of RMS, as shown in Figure 1.The graph shows that the first 700 sample groups are basically in a relatively stable state with little change, but after 700 sample groups, the curve shows obvious change. Therefore, the data group with normal sample size of the first 700 cycles and the data group with abnormal sample size of 920-953 cycles are not used temporarily because the RMS amplitude of 953-984 groups is too high.

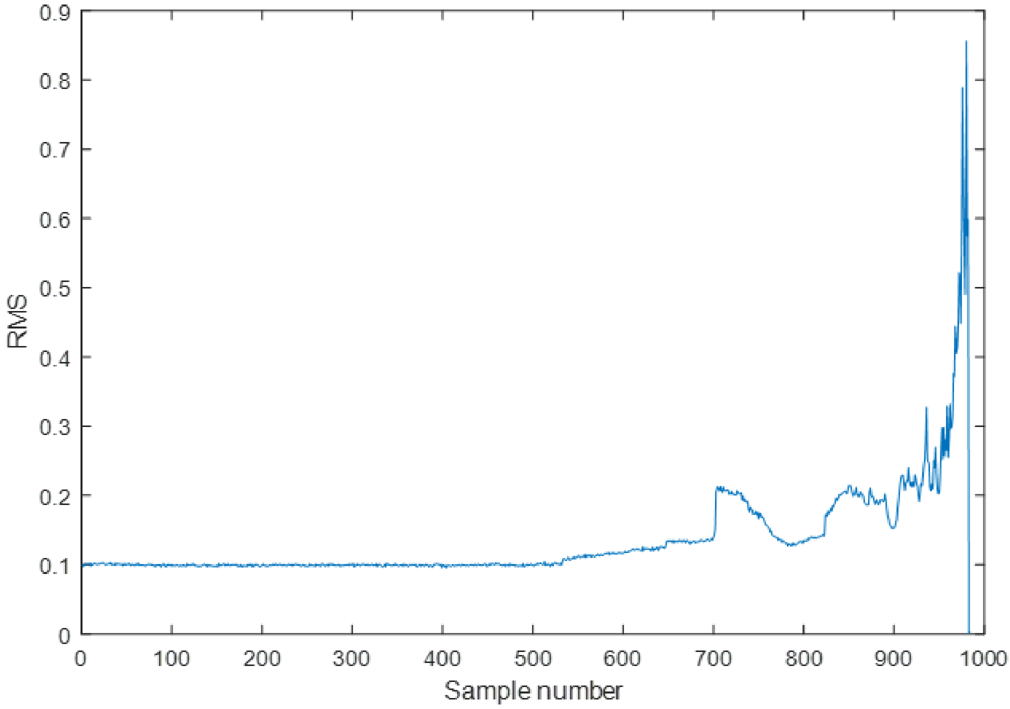


Figure 1. RMS variation trend of bearing complete life cycle.

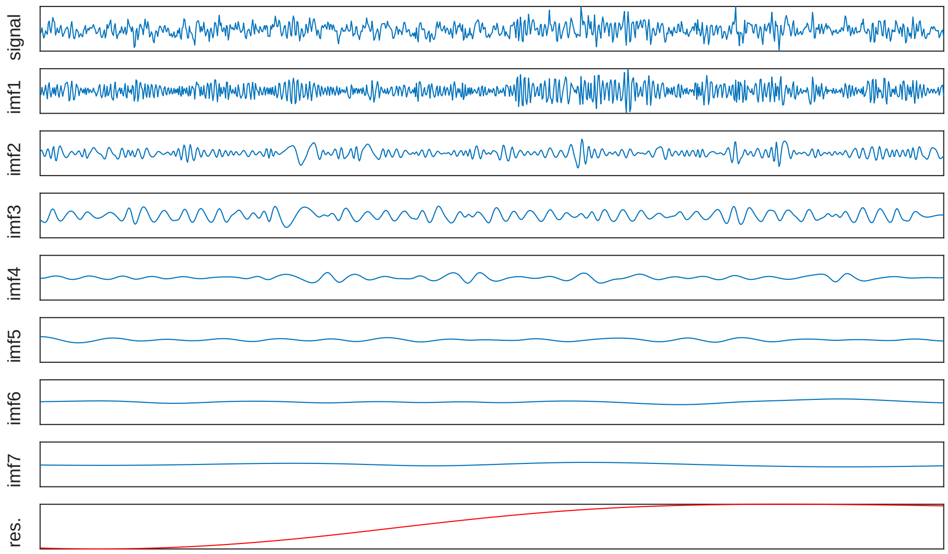


Figure 2. Signal diagram of EMD.

Since multiple IMF are obtained after empirical mode decomposition, and each IMF is different, it is found that the vibration amplitude of IMF5 is very small by observing EMD decomposition, as shown in Figure 2.Select the first 4 IMF components from the multi-layer IMF obtained by EMD decomposition, and find their corresponding mean, variance, skewness and kurtosis. Thus, the 16 dimensional characteristic matrix is formed, which constitutes the original feature space.

The validity of benchmark space and 16 dimensional characteristic matrix of normal samples are verified, as shown in figures 3 and 4.The validity of benchmark space and 16 dimensional characteristic matrix of normal samples are verified, as shown in figures 3 and 4.

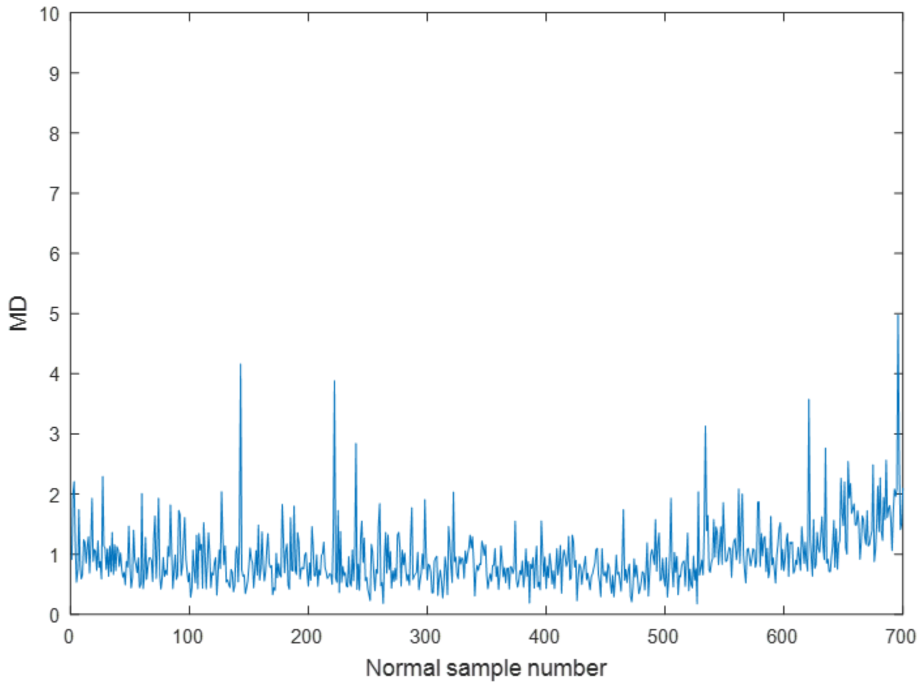


Figure 3. Mahalanobis distance of normal samples.

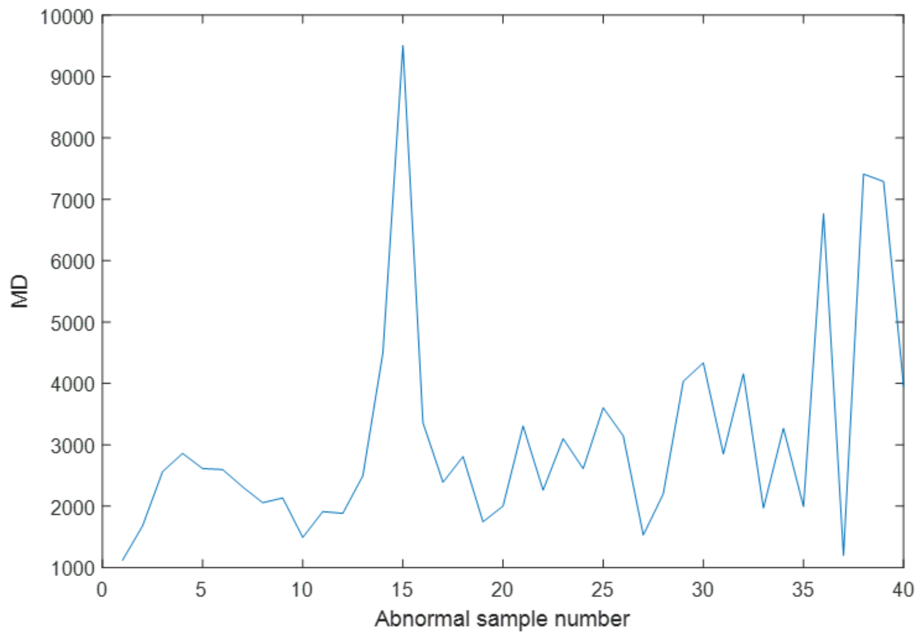


Figure 4. Mahalanobis distance of abnormal samples.

In order to optimize the reference space, the orthogonal table and SNR are replaced by the rough set optimization reference space. Firstly, the data is discretized by Boolean reasoning method, and the genetic algorithm is selected in the reduction algorithm. The 16 initial feature variables are screened by rough set method, and 7 feature variables are obtained, which are x1, x4, x5, x7, x8, x12, and x13. Therefore, these 7 characteristic variables reconstruct the reference space.

Table.1. Rough set feature selection.

project	Screening quantity	Filter results
Characteristic variable	7	x1, x4, x5, x7, x8, x12, x13

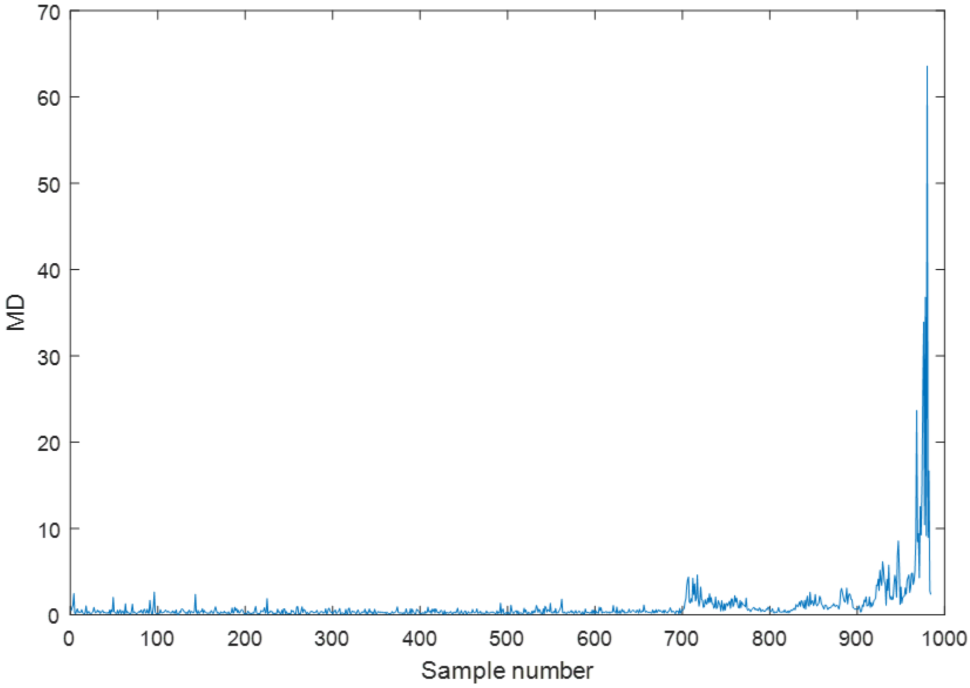


Figure 5. Mahalanobis distance throughout the life cycle

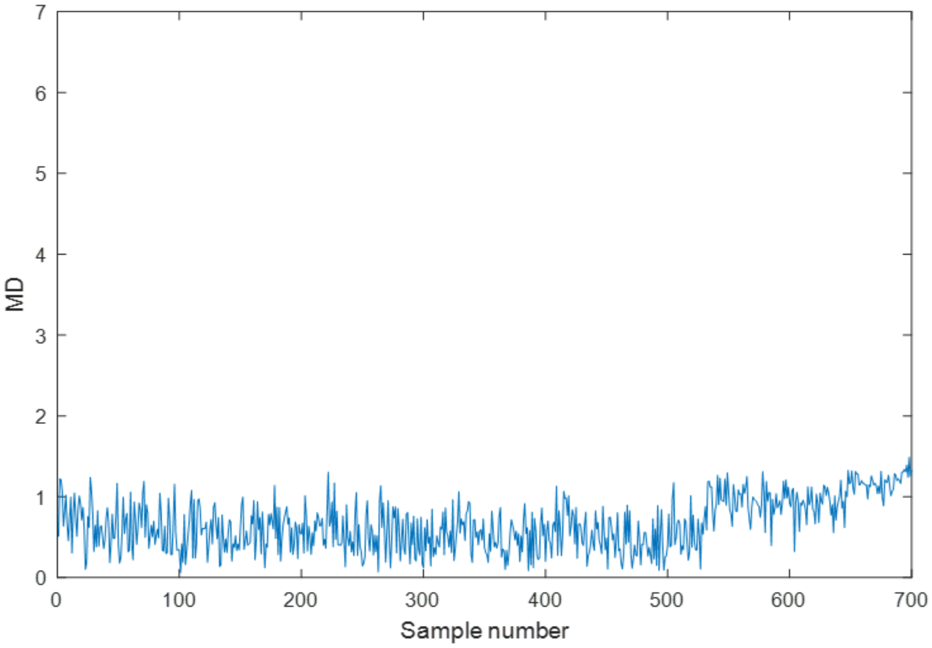


Figure 6. Mahalanobis distance after feature reduction

All data files are taken as test samples, and the Mahalanobis distance between the sample to be tested and the reduced benchmark space is calculated. From Figure 5, we can see the change of Mahalanobis distance in the complete life cycle of the reference space after reduction. When the bearing is damaged, the Mahalanobis distance is extremely large. Figure 6 shows the MD of the normal group in the optimized reference space. It can be observed from the figure that, except for individual samples, the Mahalanobis distance is basically around 1 before 530 group, and it increases significantly after 530 group. In order to further understand the trend of rolling bearing, it is necessary to carry out the next step of research.

Using the weighted Mahalanobis distance model of random forest, the importance of each channel of the characteristic variable is obtained through the random forest, and the importance of each channel is the weight of the weighted Mahalanobis distance, as shown in Figure 7.

Table 2. Importance of characteristic variables.

Characteristic variable	x1	x2	x3	x4	x5	x6	x7
Channel importance	0.388257	0.126707	0.267308	0.021806	0.047043	0.017212	0.131666

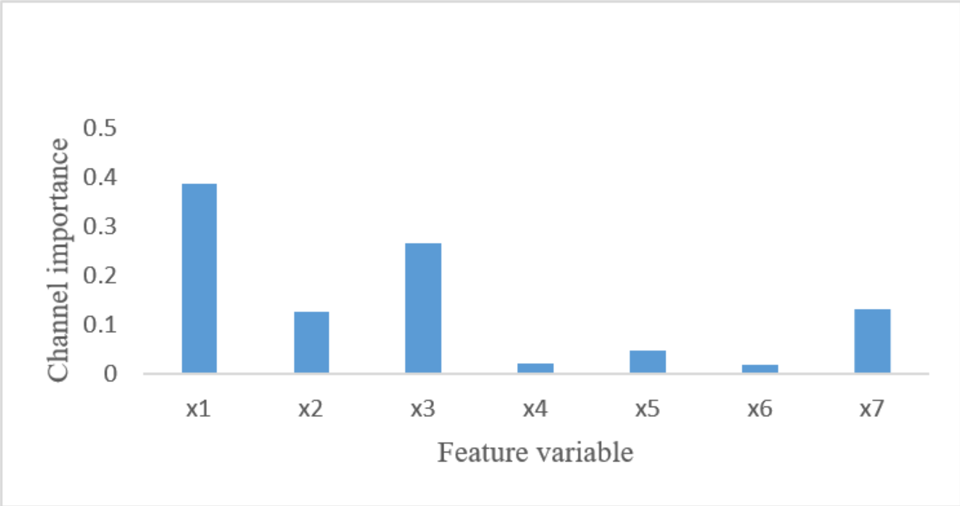


Figure 7. Importance of feature variable channels.

Calculated by formulas 5 to 6. Among them, the adjustment factor is 0.9.Based on the health assessment example of EMD-IMTS, the weighted Mahalanobis distance is obtained by using the optimized reference space, and the health index model is constructed to obtain the health assessment index of rolling bearing based on EMD decomposition of the original signal. The change trend of the value with time is shown in Figure 8.

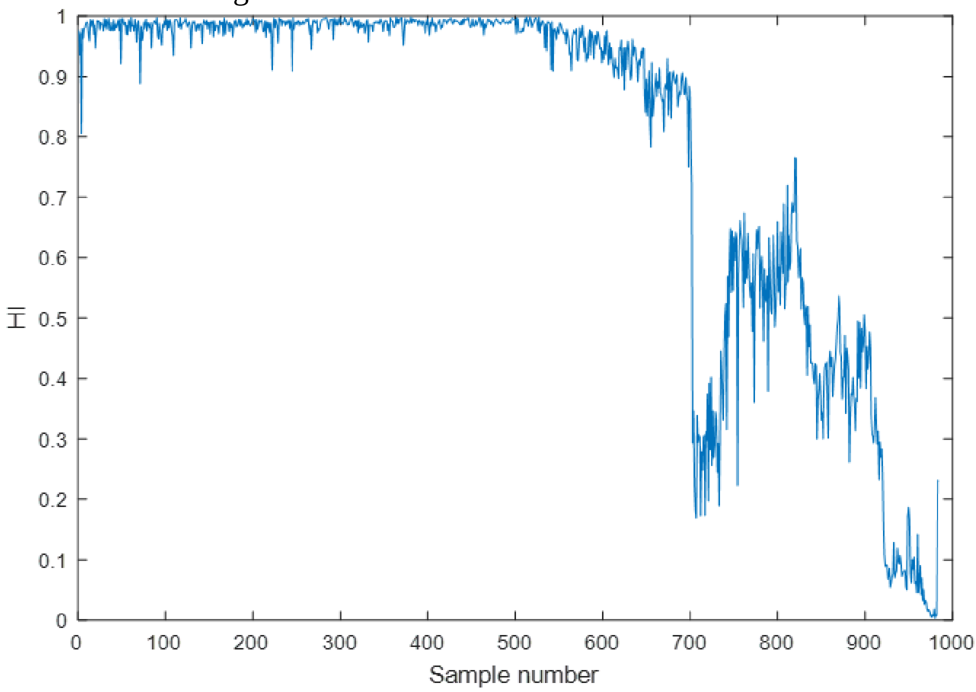
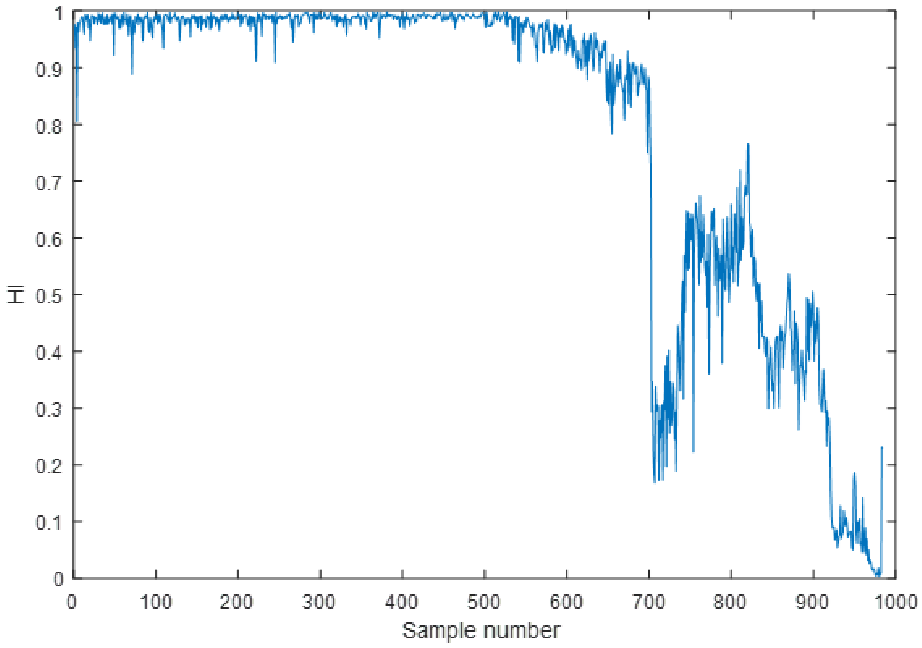
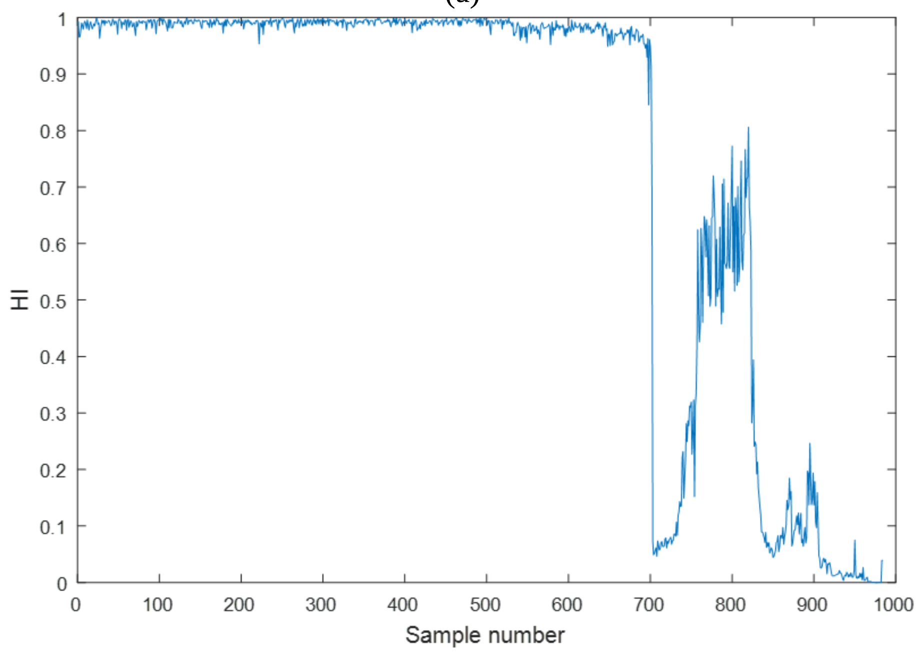


Figure 8. HI_1 of a complete bearing cycle.

Figure 8 shows the degradation process of the rolling bearing's full life cycle. Among them, the first 530 groups are in normal operation, the 530-700 groups are the degraded vibration data during normal operation, and the 700 groups have a cliff-like decline, which can be judged. The failure caused a large change in data that exceeded the normal degradation trend. The 701-900 sets of data showed the decay process of rolling bearings in the failure mode, irregular and a large range of data variation. The 900 sets reached the failure threshold and the bearing lost its function and required replacement parts. The trend change is shown in Figure 8. The state of health index can effectively reflect the operating status of the bearing. It can be applied to the initial failure monitoring and can provide maintenance and maintenance for the equipment in a timely manner, thereby saving costs and extending the life of mechanical equipment.



(a)



(b)

Figure 9. HI Comparison of bearing complete cycle (a) (b).

In order to make a comparative study and calculate the rolling bearing health assessment based on EMD-MTS, the rest of the model is unchanged, and the IMTS method is replaced by MTS method, and the health index is HI_2.After the 530-700 group, HI_1 began to show a downward trend along with the vibration, and the rolling bearing showed an early decline. HI_2 The initial decline was not obvious. In contrast, HI_1 was more significant in the early decline process. The health index of group 700 of rolling bearing decreased precipitously, the faults of group 700-900 gradually deepened, and it was in the warning stage of health state. In this stage, compared with group HI_2, the vibration reciprocating was more, the evaluation stability was poor and the convexity was obvious, and the evaluation index values had a certain overlap range. In the identification process, the warning stage might be identified as the sub-health stage. After 900 group, the rolling bearing degenerated seriously and was in the failure stage. HI_1 In 900 group declined obviously until the health index was 0. It can be seen from the above that the stability and accuracy of the health index HI_1 obtained in the health state index model proposed in this paper is better than that of HI_2, and it has higher effectiveness.

4. Conclusion

This paper proposes a health evaluation model based on EMD-IMTS, and the health index is HI_1 respectively, which can accurately reflect the health status of rolling bearings. The conclusion of this article is as follows:

(1)In order to improve the accuracy of rolling bearing health evaluation, this paper uses EMD to screen out IMF components, calculates the time-domain features of IMF components, and extracts the dimensionless and dimensionless features in time domain to form the initial feature space, which is the input space set of health state recognition method.

(2)In view of the shortcomings of MTS, rough set theory is used instead of the orthogonal table and SNR ratio to screen more effective feature variables. The weighted Mahalanobis distance is the calculation of the contribution of different feature variables. Since the random forest method can objectively assign weights to unbalanced data, the random forest is used to weight the Mahalanobis distance.

(3)A health assessment index model is proposed, and the health index HI_1 is obtained. In order to observe the change process of rolling bearing in the whole life cycle. Compare the results of the health status index of the MTS and the IMTS in this model, and use examples to prove the effectiveness and accuracy of the method.

References

[1] Chen Jing, Peng Yu, Li Qi, et al. Data-driven PHM software system for airborne equipment [J]. Journal of Electronic Measurements and Instrumentation, 2015, 29(10): 1536-1543. (In Chinese)

[2] Liao H, Shen N, Wang Y. Design and realization of an efficient environmental assessment method for 3R systems: a case study on engine remanufacturing[J]. International Journal of Production Research, 2019, 58(19): 1-24.

[3] Li X, Jiang H, Xiong X, et al. rolling bearing health prognosis using a modified health index based hierarchical gated recurrent unit network [J]. Mechanism and Machine Theory, 2019, 133: 229-249.

[4] Jiang H, Chen J, Dong G, et al. An intelligent performance degradation assessment method for bearings [J]. Journal of Vibration & Control, 2016, 23(18): 3023-3040.

[5] Zhang G, Tian F Q, She B, et al. Health indicator construction method based on the information entropy of a specific frequency band and the RBM[J]. Journal of Vibration and Shock, 2020, 39(6): 147-153+197. (in Chinese)

[6] Xiao W B, Cheney J, Zhou Y, et al. Wavelet packet transform and hidden Markov model based bearing performance degradation assessment [J]. Journal of Vibration and Shock, 2011, 30(8): 32-35. (in Chinese)

[7] Jiang H, Shang C Y, Gao R P. Wheel /rail fault noise diagnosis method based on EMD and neural network [J]. Journal of Vibration and Shock, 2014, 33(17): 34-38. (in Chinese)

[8] Wang T, Zhang M, Yu Q, et al. Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal [J]. Journal of Applied Geophysics, 2012, 83(14): 29-34.

[9] Su C T, Hsiao Y H. An Evaluation of the Robustness of MTS for Imbalanced Data [J]. IEEE Transactions on Knowledge & Data Engineering, 2007, 19(10): 1321-1332.

[10] Peng Z M, Cheng L S, Zhan J, et al. Fault classification method for rolling bearings based on the multi-feature extraction and modified Mahalanobis-Taguchi system[J]. Journal of Vibration and shock, 2020, 39(6): 249-256. (In Chinese)

[11] Lin L, Chen Z Y. Mechanical Fault Diagnosis of High Voltage Circuit Breakers Based on Rough Set Neural Networks and Vibration Signals [J]. Transactions of China Electro technical Society, 2020, 35(S1): 277-283. (in Chinese)

[12] Chen S Q, Li X Y. Gas reservoir detection based on local frequency attributes. [J]. Oil Geophysical Prospecting, 2012, 47(5): 740-746+844+678. (In Chinese)

[13] Chang Z P, Cheng L S. Multi-Attribute decision making method Based on Mahalanobis-Taguchi system and fuzzy integral [J]. Journal of Industrial Engineering/

[14] Engineering Management, 2015, 29(3): 107-115. (In Chinese)

[15] Su C T, Hsiao Y H. Multiclass MTS for simultaneous feature selection and classification [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(2): 192-205.

[16] Lai C G, Chen X H, Chen X Y, et al. A fuzzy comprehensive evaluation model for flood risk based on the combination weight of game theory [J].Nat Hazards, 2015 (77): 1243-1259.

[17] Yu J, Fang L, Cang D B, et al. Evaluation of land eco-security in Wanjiang district base on entropy weight and matter element model [J]. Transactions of the Chinese Society of Agricultural Engineering (Transaction of the CASE), 2012, 28(5): 260-266. (in Chinese)