Intelligent Diagnosis Algorithm for Bearing Faults Based on Vibration Potential Features and Affinity Propagation Clustering

Meng Li\textsuperscript{1,a}, Yanxue Wang\textsuperscript{1,2,b,*}

\textsuperscript{1}School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China

\textsuperscript{2}Beijing Key Laboratory of Performance Guarantee on Urban Rail Transit Vehicles, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

\textsuperscript{a}email: 2108230419001@stu.bucea.edu.cn

\textsuperscript{b}email: wangyanxue@bucea.edu.cn

Abstract. In order to achieve intelligent classification of bearing faults, after comparing a large number of mechanical fault signal features, this paper proposes a bearing intelligent diagnosis algorithm based on vibration potential energy feature extraction and AP clustering. The potential energy features are extracted from the multidimensional eigenmode function (IMF) of the vibration signal after the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), and after self-weighted feature selection, the Affinity Propagation Clustering (AP) algorithm is used to achieve accurate classification of unlabeled faulty bearings. After data validation, the method can be better applied to the fault classification of rolling bearings, and the performance is better than the traditional classification and diagnosis algorithm. The algorithm is used to guide intelligent fault diagnosis of unlabeled data and to improve the applicability of AP clustering algorithm in rotating machinery fault diagnosis.

1. Introduction

As the precision of machinery and equipment and production costs increase, although the incidence of mechanical failure is low, the equipment is in normal operation for a long time, but the occurrence of mechanical failure often has a lag, mechanical failure, especially rotating machinery failure caused by the economic cost and time cost loss also increases, high-speed rotating precision machinery fault prediction technology difficulty requirements and maintenance costs are increasing, the use of modern testing technology to determine the classification of mechanical failure has become an important research direction.

A large number of studies have shown that for the operational fault diagnosis of rotating machinery and equipment, the analysis of the vibration signal generated during the operation of the equipment is a practical analysis method. Since the bearing signal has the characteristics of unsteadiness and nonlinearity, related scholars have conducted a lot of research on it and obtained a lot of research results on signal processing methods. From Fourier transform to wavelet transform, from EMD (Empirical Modal Decomposition) \cite{1} to VMD (Variational Modal Decomposition) \cite{2}, research scholars are constantly improving the algorithms to obtain better signal processing results. In order to
achieve better energy feature aggregation, an improved algorithm based on simultaneous squeeze transform was proposed in [3] to achieve better signal time-frequency analysis processing. The CEEMDAN algorithm [4] is based on the Ensemble Empirical Modal Decomposition (EEMD), and overcomes the problem of EEMD losing completeness after adding white noise by adaptively adding white noise. And in the field of feature classification, the affinity propagation (AP) algorithm [5] has been receiving attention from a large number of researchers since it was proposed by Frey et al. in 2007 due to its unique clustering principle [6]. Due to its robustness, a large number of enhanced algorithms based on AP have been proposed continuously, such as hierarchical AP [7], semi-supervised AP [8], etc.

In order to improve the quality of mechanical equipment fault signal diagnosis classification, this paper proposes an intelligent diagnosis algorithm for mechanical faults based on signal vibration potential energy characteristics of tight-neighbor propagation clustering. The main innovations of the algorithm in this paper are summarized as follows.

1. Due to the large difference in distribution between different bearing data, this paper creatively proposes a signal potential energy feature extraction algorithm based on fluctuating potential energy based on the time-domain waveform characteristics of vibration signals, which is used to establish the classification feature matrix after signal decomposition.

2. The sensitive feature adaptive selection algorithm is proposed, which features that the target domain can select high weight features, i.e., the best migration features, on its own, and the feature selection can select the best migration features on its own according to the different data sets.

3. For intelligent classification of unlabeled fault signals in small samples, a clustering algorithm based on potential energy features with immediate neighborhood propagation is proposed which consists of three main stages. In the first stage, the signal is firstly decomposed by CEEMDAN and potential features are extracted; then in the second stage, the best classification features are selected according to the feature distribution; in the third stage, the best feature values are input into the migrated tight-neighbor propagation clustering algorithm to obtain the classification discriminative results of the data.

2. Algorithm Principle

In this paper, a tight-neighbor propagation clustering algorithm based on vibration potential energy features is proposed and designed for fault classification in the field of mechanical fault diagnosis. The designed network structure is shown in Fig 1. It includes a potential energy feature extractor, a feature self-weighting analysis algorithm and an AP clustering category classifier, as shown in Fig. 1 which illustrates the analysis process of the algorithm in this paper.

2.1. Potential energy feature extraction algorithm

Based on the energy perspective analysis of the signal's waveform, a potential energy feature extraction algorithm of the fault signal is proposed by taking the amplitude change of the signal as the calculation criterion of the potential energy change. The algorithm design is based on the multi-dimensional eigenmode function (IMF) after CEEMDAN decomposition to further extract the
potential energy characteristics of the signal. The potential energy characteristics of the signal are obtained by calculating the modal components of the original signal decomposition. The final decomposition of the algorithm decomposes the original signal into:

$$x(n) = \sum_{k=1}^{K} IMF_k + R(n)$$  \hspace{1cm} (1)

where $IMF_k$ is the $k$ modal components obtained from the decomposition and $R(n)$ is the residual of the decomposition.

Based on the energy perspective analysis of the signal time domain waveform, the amount of potential energy change of the signal is taken as the potential energy feature of the signal, and the vibration potential energy calculation algorithm of the signal is proposed to extract the energy feature of the $k$ modal components obtained from the CEEMDAN decomposition as follows:

$$E_p^k = \sum_{i=1}^{n-1} [x_i(i+1) - x_i(i)]^2, \quad k = 1, \ldots, K$$  \hspace{1cm} (2)

where $x_i(i)$ is the element in the $k$ modal component $IMF_k$.

Using the above algorithm, The $K$ potential features of the original signal $x(n)$ can be extracted and used as the classification discriminant features of the signal.

2.2. Feature self-weighting analysis algorithm

The feature self-weight analysis algorithm proposed in this paper is an algorithm that is applicable to mechanical fault analysis and can calculate the weight value by the features themselves after analyzing a large amount of mechanical fault data. The algorithm relies only on the weight analysis of features without category labels, thus reducing the input volume of the algorithm. To improve the classification effect, the feature set is first normalized.

$$v_{j,k} = \sqrt{v_{j,k}^2}$$  \hspace{1cm} (3)

Then calculate the self-similarity coefficient:

$$SS_{nj,k} = \left\| v_{n,k} - v_{j,k} \right\|^2, \quad n, j = 1, 2 \cdots J; k = 1, 2 \cdots K$$  \hspace{1cm} (4)

The self-similarity coefficient represents a measure of the variability between the same features for different samples. The weight size of each feature in the dataset is obtained according to the self-weighting algorithm, and the sensitive features with high weights among them are selected as classification migration features, and these classification features form the similarity matrix of the sample set as the initial input for clustering.

Some text.

2.3. Affinity Propagation Clustering

The principle of the AP clustering algorithm is in taking each sample point of the algorithm input as a potential cluster center representative point, and the goal is to find the appropriate cluster representative point so that the clustering energy is minimized, where the measure of clustering energy is numerically characterized using negative Euclidean distance, and the objective function of AP clustering is obtained as:

$$\max S(c) = \sum_{i=1}^{N} S(i, c_i) + \sum_{i=1}^{N} \delta_k(c)$$  \hspace{1cm} (5)
where \( S(i, c_j) \) denotes the distance from the \( i \) point to the \( c_j \) potential representative point, \( \delta_i(c) \) denotes the penalty to the potential cluster center representative point, and the penalty function is defined as:

\[
\delta_i(c) = \begin{cases} 
-\infty, & \text{if } c_i \neq k \text{ but } \exists i : c_i = k \\
0, & \text{otherwise}
\end{cases}
\]  

(6)

If the data point \( c_i \) chooses \( k \) as its class representative point, i.e. \( c_i = k \), then the data point \( k \) must choose itself as the class representative point, i.e. \( c_k = k \).

When the amount of data is sufficient, the AP algorithm can accurately find the best clustering representative points and category assignment results without the influence of external factors.

3. Experimental Validation

3.1. Experimental Validation of MFPT Data

This set of experimental validation selected three different states of bearings from the Mechanical Failure Prevention Technology (MFPT) bearing fault diagnosis dataset, namely normal, outer failure and inner failure, with 50 sets of samples each, and tested the accuracy of data classification by using different feature extraction algorithms and different classification algorithms. The comparison methods and classification effects used for data validation are shown in Table 1 as follows.

| Group | Signal decomposition method | Feature       | Clustering Methods | K-means Clustering accuracy | Number of AP Clustering categories | AP Clustering accuracy |
|-------|-----------------------------|---------------|--------------------|-----------------------------|----------------------------------|-----------------------|
| G1    | EEMD                        | Energy        |                   | 62%                         | 3                                | 94.5%                 |
| G2    | EEMD                        | Potential     |                   | 59.5%                       | 3                                | 66.5%                 |
| G3    | CEEMDAN                     | Energy        |                   | 62.5%                       | 3                                | 97.5%                 |
| G4    | CEEMDAN                     | Potential     |                   | 61%                         | 3                                | 99.5%                 |

In this set of experimental validation eight sets of experimental comparison experiments were used to verify the effectiveness of the algorithm, in which the signal decomposition methods were EEMD and CEEMDAN, the feature extraction was the traditional time-frequency domain energy features and the potential energy features proposed in this paper, and the classification methods were K-means clustering and AP clustering. Firstly, the original feature set obtained by using four different feature extraction methods was analyzed for feature self-weighting. Sensitive features with high weights were selected according to the weight analysis to form a sensitive feature set, which were input into the K-means clustering and AP clustering algorithms, respectively, and the clustering results were obtained as shown in Fig 2-Fig 9.

![Fig 2 G1 Cluster(K-means)](image)
![Fig 3 G1 Cluster (AP)](image)
![Fig 4 G2 Cluster(K-means)](image)
3.2. MFS-MG experimental data validation

This set of experimental validation selected four different states of bearings from the bearing failure dataset of MFS-MG experimental bench of Guilin University of Electronic Science and Technology, which are normal, outer ring failure, inner ring failure and rolling body failure, 50 groups of samples each, and tested the accuracy of data classification by using different feature extraction algorithms and different classification algorithms. The comparison methods and classification effects used for data validation are shown in Table 2 as follows.

| Group | Signal decomposition method | Feature   | Clustering Methods | K-means Clustering accuracy | Number of AP Clustering categories | AP Clustering accuracy |
|-------|-----------------------------|-----------|--------------------|-----------------------------|-----------------------------------|-----------------------|
| G5    | EEMD                        | Energy    |                   | 84.5%                       | 4                                 | 84.5%                 |
| G6    | EEMD                        | Potential energy |               | 34.5%                       | 5                                 | 38%                   |
| G7    | CEEMDAN                     | Energy    |                   | 90%                         | 4                                 | 96.5%                 |
| G8    | CEEMDAN                     | Potential energy |               | 89.5%                       | 4                                 | 100%                  |

The classification results of the algorithm validated with this dataset are shown in Fig 10-Fig17.
3.3. Analysis of experimental results

After decomposing the three stages of the algorithm to set up a control group for mixed comparison, it is found that the algorithm proposed in this paper shows better classification performance than other algorithms under different datasets. In the experimental results, it can be seen that the classification effect reaches 99.5% in the data set of MFPT and 100% in the data set of MFS-MG experimental bench, and the results show that the algorithm can be effectively applied in the detection and classification of bearing faults and can achieve good results.

4. Conclusion

In this study, after conducting a large number of existing research deficiencies, the proposed potential energy feature extraction algorithm, as a new feature extraction method, achieves better results in the field of mechanical fault diagnosis, and combined with the newly proposed self-weight analysis algorithm and AP clustering, can be effectively applied to the detection and classification of mechanical bearing faults, and can reach 99.5% or even 100% in classification accuracy. Compared with the EEMD processing and traditional time-frequency domain energy extraction algorithm, it has outstanding performance and can be effectively applied to the classification of bearing faults.

References

[1] Alsalah Ali, Holloway Damien & Lavroff Jason(2021). Identification of wave impacts and separation of responses using EMD. 151

[2] D R AGOMI R ETSKIV K. ZOSSO D. Variational mode decomposition [ J ] . IEEE Transactions on Signal Processing, 2014, 62(3) : 531-544

[3] SUN Lingfang, XU Manfei, PIAO Heng, et al. Self-adaptive noise denoising for ultrasonic detection signal based on improved CEEMD [ J ] . Journal of Vibration and Shock, 2017, 36(Liu, Wang, & Xu, 2020)( 20): 225—232.

[4] Liu, Q., Wang, Y., & Xu, Y. (2020). Synchrosqueezing extracting transform and its application in bearing fault diagnosis under non-stationary conditions. Measurement, 108569. Retrieved from http://www.sciencedirect.com/science/article/pii/S0263224120310915. doi:https://doi.org/10.1016/j.measurement.2020.108569

[5] Hang, W., Chung, F. L., & Wang, S. (2016). Transfer affinity propagation-based
clustering. *Information ences*, 348, 337-356.

[6] Frey BJ, Dueck D. Clustering by passing messages between data points. Science, 2007,315(5814):972–976. [doi: 10.1126/science.1136800]

[7] Xiao J, Wang J, Tan P, Quan L. Joint affinity propagation for multiple view segmentation. In: Proc. of the 11th IEEE Int’l Conf. on Computer Vision. IEEE Press, 2007. 17. [doi: 10.1109/ICCV.2007.4408928]

[8] Xiao Y, Yu J. Semi-Supervised clustering based on affinity propagation algorithm. Ruan Jian Xue Bao/Journal of Software, 2008,

[9] 19(11):2803 2813 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/19/2803.htm [doi: 10.3724/SP.J.1001.2008.02803]