COMPOSITE FAULT DIAGNOSIS IN ROTATING MACHINERY
BASED ON MULTI-FEATURE FUSION

Summary

The rotating machinery working in complex environments of petrochemical units often develops composite faults and its vibration signal exhibits multicoupling, fuzziness, and nonlinearity, making it difficult to effectively diagnose composite faults. This paper proposes a composite fault diagnosis for rotating machinery based on multi-feature fusion. This method firstly extracts the time domain, the frequency domain and the dimensionless feature information, using the correlation analysis and normalization to obtain bodies of evidence with different features. Then, according to the fusion rules of the evidence theory, the synthesis of different bodies of evidence is completed. Finally, the feasibility of the proposed method is verified. The experimental results show that the accuracy of the proposed method exceeds 90%, thus it has been shown that the composite fault diagnosis of rotating machinery in petrochemical units is effective.

Key words: composite fault diagnosis, time domain, frequency domain, high-value dimensionless, feature fusion

1. Introduction

Numerous scholars and engineers have devoted a great deal of time and effort to the study of composite fault diagnosis methods. These methods focus primarily on the analysis and identification of time-domain signals, frequency-domain signals, and feature fusion in order to diagnose equipment failure.

The time-domain method analyses the differential equation or transfer function of a signal. The signal changes over time and under the influence of external factors. There is an intuitive and precise feature that can directly reflect the operating status of the operating system of a unit [1-2]. Li et al. explained the frequency characteristics of a stationary shaft gearbox under different fault conditions using dynamic differential equations and utilizing the derived modulation features as indicators for the fault diagnosis [3]. Jiang et al. employed a finite element model of a flexible shell to compute the time-varying transfer functions determining the amplitude characteristics of vibration signals. They also derived full-frequency range
vibration signals through the convolution of the excitation force and flexible transfer functions, thereby revealing fault characteristics related to the sun gear and ring gear cracking [4]. Additionally, Jiang et al. established a rigid-flexible coupled dynamics and signal convolution model, unveiling the vibration response mechanism of stationary shaft gear systems with cracks and they selected the modulation sideband features of the shell vibration response signal as indicators for crack faults [5].

Due to the increasing complexity of the speed-increasing box structure and the influence of site conditions, it is challenging to install a vibration sensor. Scholars use the changing law of the acoustic emission signal to complete the identification and diagnosis of rolling bearing faults [6-7]. The application of this law can solve the problem of difficulty in the effective evaluation of the operating status of units under variable conditions. Zhan et al. [8] considered both time-domain and frequency-domain information and used the multivariate coefficient of variation (MCV) to fuse vibration residuals and environmental parameters, creating an anomaly assessment metric. Sui et al. [9] based their study on the analysis of switch states and proposed a time-domain model for post-inverter fault currents to estimate the currents in each phase following an open-circuit switch fault.

The frequency-domain analysis technology uses the frequency feature as a sample, analyses the corresponding relationship between the system internal structure and the unit performance, and reveals the inherent characteristics of the signal [10-11]. Hao et al. [12] used the operating unit vibration signal for the bispectrum analysis and reflected the fault information implicit in the characteristic of the fault signal of a unit bearing. Feng et al. [13] took the normal state and different fault state signals of an epicyclic gearbox as the feature, established a vibration signal change model and an analytic formula of the frequency-domain signal. Zheng et al. [14] combined the square envelope analysis and introduced the Holo-Hilbert square spectrum analysis to enhance the separation efficiency of fault feature detection. Wang et al. [15] proposed a planetary gearbox fault diagnosis scheme that utilized multi-criteria fault feature selection and heterogeneous ensemble learning classification. This approach refers to the fault diagnosis on planetary gearboxes by extracting high-dimensional fault features from vibration signals in the time domain, frequency domain, and time-frequency domain.

Evidence theory is concerned with probabilities that are ambiguous and uncertain. It can be used to measure and process complex multi-source, conflicting, and uncertain information. Dempster proposed this method in 1967, and his student Shafer improved and popularised it further. Currently, this method is employed extensively in the fields of equipment fault diagnosis, expert systems, medical problem diagnosis, target tracking, and command dispatching [16-19]. In order to solve the problem of global conflicts or local conflicts between evidence, scholars made improvements in three aspects [20-22]. Firstly, the experimental method was improved by using combination rules. When there is a high degree of conflict between the evidence, the combination rule is used to combine the evidence. Secondly, considering the conflicts and uncertainties between the evidence, the probability function of the original feature was adjusted and modified. Thirdly, diagnostic results based on the global trust distribution function resulted in the modification of the research model and the way decisions, inferences or classifications are made, to determine the most likely type of fault. In the production process, these three methods can solve the problem of fusing information from multiple sources. However, the information source is primarily derived from the actual situation and site requirements. It is difficult for a combined evaluation criterion to become effective. In response to the above-mentioned problems, Smets [23] considered the incomplete identification framework and the evidence conflict problem, and proposed a transferable belief model (TBM) model. The conflict part reliability is assigned to the empty set, resolving the conflict and evaluation problems of the evidence factor. However, this method simply averages the evidence to be fused and does not consider the correlation between sample features [24]. Wang et al. [25]
used a Markov chain to model the uncertainty information in the evidence theory, explored the
ordered rules within random evidence, and obtained evidence support to reduce the impact of
high-Conflict evidence. An analysis of the research status of the fault diagnosis in petrochemical
units shows that the main problems are the following: Firstly, there are few research results for
composite fault diagnoses for petrochemical units, and the high-value sample features are not
considered comprehensively. Secondly, the composite fault diagnosis research process lacks
the feature fusion and it is difficult to effectively carry out the composite fault diagnosis.
Thirdly, in the diagnosis research process only the time-domain or frequency-domain features
are used in the composite fault diagnosis, thus it is difficult to achieve results.

The main contribution of this study refers to the following:

(1) This paper proposes a new data fusion method for composite fault diagnosis in
petrochemical units based on time domain, frequency domain and dimensionless indexes.

(2) This method applies correlation analysis and normalization processing to obtain
bodies of evidence with different features.

(3) Fusion rules that fuse different bodies of evidence to identify composite faults have
been developed.

2. Relation theory

2.1 Correlation analysis and normalisation

The time-domain signal \( x(t) \) and the frequency domain are obtained from a vibration
collector, and the composite fault dimensionless feature is obtained by comparing two
dimensional parameters. However, analysing the vibration signals of rotating machinery using
only composite fault dimensionless features is not accurate, which is defined as follows:

\[
\zeta_x = \left[ \frac{\int_{-\infty}^{\infty} |x| p(x) dx}{\int_{-\infty}^{\infty} x^l p(x) dx} \right]^{\frac{1}{l}} = \frac{\sqrt{\int E(|x|^l)}}{\sqrt{\int E(|x|^m)}} \tag{1}
\]

In the formula, \( x \) is the amplitude of the vibrational random time-domain signal, \( p(x) \) is
the probability density function of the signal \( x \), \( l \) and \( m \) are the numerator and denominator
coefficients, respectively.

The time domain, frequency domain and dimensionless feature are utilised to perform a
correlation analysis and normalisation, to construct a framework of fusion rules of the evidence
theory and to achieve the multiple feature fusion. Pearson's correlation coefficient \( [26] \) is used
to measure the linear correlation between two variables. Among them, \( \text{cov}(X, Y) \) is the
covariance of training sample \( X \) and testing sample \( Y \) in the petrochemical unit, while \( \sigma_X \) and
\( \sigma_Y \) are the standard deviations of the two samples, respectively. The definition can be expressed as:

\[
\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{2}
\]

The Pearson correlation coefficient \( r \) ranges from -1 to 1. The negative \( r \) of -1 indicates
a perfect negative correlation, \( r \) of 1 indicates a perfect positive correlation, and \( r \) of 0 indicates
no linear correlation. By estimating the covariance and standard deviation of the time-frequency
domain and high-value dimensionless sample features, the correlation coefficient between the
training sample \( X \) and the test sample \( Y \) is
\[ r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{l_{XY}}{\sqrt{l_{XX}l_{YY}}} \]  

(3)

In the formula, the sum of squared deviations of \( X \) and \( Y \) is: 
\[ l_{XX} = \sum (X - \bar{X})^2, \]
\[ l_{YY} = \sum (Y - \bar{Y})^2, \]
and the sum of deviations of \( X \) and \( Y \) is 
\[ l_{XY} = \sum (X - \bar{X})(Y - \bar{Y}). \]

The time domain, frequency domain and dimensionless feature correlation data are mapped onto the range \([0, 1]\), and the correlation coefficient is transformed into the evidence weight ratio. Suppose that there are \( n \) different bodies of evidence, the training samples \( X = \{x_1, x_2, \ldots, x_n\} \), and the weight coefficient is [27]:
\[ y_k = \frac{x_k}{\sum_{k=1}^{n} x_k} \]  

(4)

In the formula, the variable \( Y = \{y_1, y_2, \ldots, y_n\} \), the \( Y \) elements range \([0, 1]\), and 
\[ \sum_{k=1}^{n} y_k = 1. \]

2.2 Evidence theory

Evidence theory is a mathematical framework used for handling uncertainty and synthesizing information from different sources. It is clearly related to other frameworks, such as probability theory, possibility theory, and imprecise probability theory. Dempster-Shafer evidence theory is rooted in two principles: deriving the belief of a problem from subjective probabilities of relevant questions and combining the belief from independent evidence using Dempster's rule [28]. According to the time domain, frequency domain, and dimensionless features, the evidence factor is obtained through the correlation analysis and normalization processing.

Suppose that the types of faults occurring in petrochemical units are included in the identification framework \( \Theta \). The elements between different faults are independent and mutually exclusive, denoted as \( \Theta = (k_1, k_2, \ldots, k_n) \), and \( 2^\Theta \) represents the power set of the petrochemical unit identification framework. Suppose that the recognition framework for any hypothesis is \( \Theta \), and the trust degree \( Bel(A) \) is defined as the sum of the basic probabilities corresponding to all subsets in \( A \). The basic reliability function [29] is as follows:
\[ BPA(A) = \begin{cases} 0 & A = \emptyset \\ \sum_{A \subseteq 2^\Theta} m(A) = 1 & A \neq \emptyset \end{cases} \]  

(5)

In the formula, \( m \) is introduced as an evidence factor to characterize the trust in the features of the framework sample, and this trust distribution is named complete trust \( A \). It reflects the reliability of \( A \) itself, indicating that the evidence related to the feature is fully trusted, i.e. that the feature had been proven.

The identification framework [30] reflects the occurrence probability of different faults of petrochemical units. The identification framework of the known fault types of the petrochemical units is
\[ \Theta = \{\theta_1, \theta_2, \ldots, \theta_n\} \]  

(6)
Suppose that the existence of an identification framework $\Theta$ for petrochemical facilities, $2^\Theta$ is the set of all subsets of $\Theta$, and if $m$ was $2^\Theta \rightarrow [0, 1]$, the following formula should be satisfied:

$$\sum_{A \subseteq 2^\Theta} m(A) = 1, \quad m(\emptyset) = 0 \quad (7)$$

Suppose that function $Bel$ is the confidence function of the petrochemical unit identification framework $\Theta$, which represents the evidence factor trust degree of the time-frequency domain and high-value dimensionless features in $A$, satisfying the formula:

$$Bel(A) = \sum_{B \subseteq A} m_1(A_j) \quad \forall A \subseteq \Theta \quad (8)$$

In the formula, $Bel(A)$ is the lower limit estimate of the petrochemical unit reliability.

Suppose that function $PI$ is the likelihood function of the identification framework $\Theta$ of the fault type of the petrochemical unit, and function $PI$ satisfies $2^\Theta \rightarrow [0, 1]$, then there is

$$PI(A) = \sum_{B \setminus A \neq \emptyset} m_2(B_j) \quad \forall A \subseteq \Theta \quad (9)$$

In the formula, $PI(A)$ is the upper limit estimate of the petrochemical unit reliability.

3. Experimental research of the petrochemical unit composite fault

3.1 Experimental conditions and feature extraction

In this study, a platform for diagnosing faults in the monitoring equipment of a petrochemical unit, data resources and test conditions were investigated. The petrochemical unit test platform is shown in Figure 1.

![Fig. 1 Fault diagnosis platform in a petrochemical unit](image)

The main fault types investigated in the experimental study are: bearing inner ring fault, bearing outer ring fault, bearing ball fault, and large and small gear fault. The rolling bearing parameters are as follows: the outer diameter is 110 mm, the inner diameter is 50 mm, the thickness is 27 mm, the diameter and the number of rolling elements are 14 mm and 13, respectively. The number of teeth of the big gear and the small gear are 29 and 23, respectively, and the motor speed is 1,000 r/min.

This experiment selects four types of composite faults and one single fault. Time-domain and frequency-domain features from 10,000 vibration signal data are extracted for each fault. The objects of the study are: normal state (ZC) and composite faults types: gear fault and bearing inner ring fault (CNQ), gear fault and bearing outer ring fault (CWQ), gear fault and
bearing ball fault (CGZ). The composite faults were identified through the time domain, frequency domain and dimensionless feature fusion.

![Time-domain and Frequency-domain diagrams](image)

**Fig. 2** Time domain and frequency domain

![Dimensionless feature distribution](image)

**Fig. 3** Dimensionless feature distribution of composite faults

It can be seen from Figure 2 that there is a considerable overlap between the normal state and the composite fault. The distinction between different features is obvious in Figure 3. The analysis of Figures 2 and 3 shows that only a single feature cannot be used to diagnose faults, and it is challenging to differentiate between different faults. In Section 3.2, 1,000 sets of dimensionless features were extracted from the data set, with 70% of features being utilized as testing features and 30% as training features. This section also demonstrates the accuracy of using the time-domain feature evidence factor ($m_1$), the frequency-domain feature evidence factor ($m_2$), the high-value dimensionless feature evidence factor ($m_3$), and other algorithms to identify composite faults, and compares the accuracy with the data fusion method for composite fault recognition proposed in this paper.

### 3.2 Experimental verification of the method

The diagnosis results for five composite fault types, i.e. the gear fault and bearing inner ring fault (CNQ), the gear fault and bearing outer ring fault (CWQ), the gear fault and bearing ball fault (CGZ), the large and small gear fault (CQC), and the bearing ball fault (Q) are shown in Figure 4.
In the case of the CNQ fault diagnosis, the evidence factor \( m_1 \) for the time-domain diagnostic method is 0.3125, as shown in Figure 4. In the set of five major fault types, the ZC axis has the maximum value, resulting in the time-domain diagnostic result being ZC. However, the actual fault in the equipment is CNQ, illustrating that the time-domain diagnosis method cannot accurately determine the fault type. On the other hand, the evidence factor fusion result for the fusion method is 0.4956, the maximum value is on the CNQ axis in the two-dimensional plot and is higher than the other three fusion methods. Taking the CNQ diagnosis analysis as example, the multi-feature fusion method based on the time domain, frequency domain and dimensionless feature to accurately diagnose the fault types. This provides a new option for the composite fault detection on the monitoring equipment platform of a petrochemical unit.

3.3 Fault diagnosis results for different methods

In the preceding section, the accuracy and superiority of the diagnostic method based on the time-domain, frequency-domain, and dimensionless indices were validated. The method presented in this paper was also compared with six other methods, i.e. the naive Bayes model (NBM), quadratic discriminant analysis (QDA), support vector machine (SVM), k-nearest neighbours algorithm (KNN), deep neural network classifier (DNC), and convolutional neural network (CNN). The comparative results are presented in Figure 5.
Figure 5 illustrates the diagnosis accuracy of the method presented in this paper and the accuracy of different algorithms in identifying composite faults using the time-domain features, the frequency-domain features, and the dimensionless index. The comparative results reveal that the diagnosis accuracy of the data fusion method for the composite fault diagnosis in rotating machinery based on the multi-feature fusion presented in this study is 91.5%, 89% and 89.5% for CGZ, CNQ, and CWQ, respectively. These values outperform those of the six different algorithms (NBM, QDA, SVM, KNN, DNC, CNN), highlighting the superiority of our method in the area of composite fault diagnosis.

4. Conclusion

In this paper, the evidence theory fusion rules combining bodies of evidence as time domain, frequency domain, and dimensionless features were applied to diagnose composite faults in rotating machinery. Five typical fault samples (CNQ, CWQ, CGZ, CQ and Q) were used to diagnose five major composite fault types. The experimental results show that this method has high accuracy in the composite fault diagnosis. When compared with six different algorithms (NBM, QDA, SVM, KNN, DNC, CNN), it is proven that the accuracy of the proposed method exceeds the accuracy of other methods.

Acknowledgments

The research was partially supported by the Key Project of Natural Science Foundation of China (Grant No. 61933013) and the National Nature Science Foundation of China (Grant Nos. 61673127, 61973094). The study presented in this paper was also partially supported by the Natural Science Foundation of Guangdong (2022A1515010599), Maoming Green Chemical
REFERENCES


[6] Shao L Q; Zhang Q H; Lei G W; Su N Q; Yuan P H. An Immune Detector-Based Method For the Diagnosis Of Compound Faults In a Petrochemical Plant. Transactions Of Famen, 2022,03:11-12. https://doi.org/10.21278/463033721


Submitted: 13.5.2023
Accepted: 25.10.2023

Nai-quan Su*
Guangdong Provincial Key Lab of Petrochemical Equipment and Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China.
High-Tech Institute of Xi’an, Xi’an 710025, China
Qing-hua Zhang
Yi-dian Chen
Xiao-xiao Chang
Yang Liu
Guangdong Provincial Key Lab of Petrochemical Equipment and Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China
*Corresponding author:
534306286@qq.com