

Research Article

A Diagnosis Method for Rotation Machinery Faults Based on Dimensionless Indexes Combined with K -Nearest Neighbor Algorithm

Jianbin Xiong,^{1,2} Qinghua Zhang,^{1,2} Zhiping Peng,² Guoxi Sun,²
Weichao Xu,³ and Qi Wang²

¹Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis, Maoming 525000, China

²School of Computer and Electronic Information, Guangdong University of Petrochemical Technology, Maoming 525000, China

³School of Automation, Guangdong University of Technology, Guangzhou 510006, China

Correspondence should be addressed to Guoxi Sun; 158011382@qq.com

Received 30 October 2014; Revised 3 February 2015; Accepted 3 February 2015

Academic Editor: Gang Li

Copyright © 2015 Jianbin Xiong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is difficult to well distinguish the dimensionless indexes between normal petrochemical rotating machinery equipment and those with complex faults. When the conflict of evidence is too big, it will result in uncertainty of diagnosis. This paper presents a diagnosis method for rotation machinery fault based on dimensionless indexes combined with K -nearest neighbor (KNN) algorithm. This method uses a KNN algorithm and an evidence fusion theoretical formula to process fuzzy data, incomplete data, and accurate data. This method can transfer the signals from the petrochemical rotating machinery sensors to the reliability manners using dimensionless indexes and KNN algorithm. The input information is further integrated by an evidence synthesis formula to get the final data. The type of fault will be decided based on these data. The experimental results show that the proposed method can integrate data to provide a more reliable and reasonable result, thereby reducing the decision risk.

1. Introduction

Large rotating machinery and equipment (such as steam turbines, rotary bearings, fans, and compressors) are key equipment in petroleum, chemical, metallurgy, machinery manufacturing, aerospace, and other important engineering fields. Such equipment requires high safety and reliability [1], so the study of fault diagnosis methods for these types of equipment has been a hot topic in this field. Vibration monitoring signals have lots of nonlinear, random, and nonergodic information, which causes many complications in the fault signal analysis when rotating machinery does not work [2]. The time-domain signal of vibration is the most basic and original signal. References [1, 2] extracted failure characteristics directly from the time-domain signal and analyzed the fault diagnosis, showing that maintaining the basic characteristics of the signal will be very beneficial. References [2–4] used the probability density function of the vibration signal to derive

dimensional indexes (the mean and root mean square values, etc.) and dimensionless indexes (waveform, margin index, pulse, etc.) in the amplitude domain. In practice, although a dimensional index is sensitive to the fault characteristics, its value will increase with the development of the fault. In addition, as the working conditions (load, speed, etc.) change, these are easily affected by interference, which reduces their performance as diagnostic measures [3]. By contrast, the dimensionless indexes are not sensitive to the disturbance of the vibration monitoring signal and the performance is stable. In particular, these dimensionless indexes are sensitive to neither the changes in amplitude nor the frequency of the signal. That is, they have little relationship to the particular working conditions of the machine [1–3]. Dimensionless indexes have been widely used in the fault diagnosis of rotating machinery. For the dimensionless indexes, pulse index and kurtosis index are more sensitive to impact type fault, especially in the early fault, the large amplitude of

the pulse is less. Compare to other parameters kurtosis index and pulse index rise faster, so that fault of the range is larger, it is difficult to determine fault type [3, 4].

Solving the above problems requires the use of an effective method which can process uncertain information rationally, systematically, and flexibly [5]. Evidence theory can effectively express and deal with uncertain and imprecise information and other problems [6]. However, in the actual information fusion system, the interference due to the natural environment or human disturbances often leads to conflicting reports by the sensors [7]. Traditional D-S evidence theory cannot effectively deal with the integration of conflicting information. When evidence conflict exists between global or local information, using D-S combination rules for fusion leads to fusion results which are contrary to the real value [1, 8]. Therefore, when a high degree of conflict exists between the evidence, achieving effective integration between the evidence is an urgent problem that needs to be solved. At present, there are three major ways for treating evidence conflict [7]: (1) in the improvement method based on D-S combination rules, there is a high degree of conflict between the evidence, and the D-S combination rules for fusion are used directly. This often produces unreasonable fusion results. This kind of methods has been presented by Yager [9], Dubois and Prade [10], and Feng et al. [7]. (2) In the second method, which is based on modifying the original sources of evidence [8, 11], the conflict evidence is preprocessed first, and then evidence combination rules are used for fusing evidence. (3) The third method is to modify the model from the sources based on a known model.

These three methods can solve some practical application problems. Which method is used depends on the actual situation and the need. In the absence of an effective combination rule evaluation criteria, it is difficult to determine which combination rule is "the most excellent." Smets et al. put forward a TBM model and believed that the reliability of the evidence conflict which resulted from incompleteness of recognition framework should be assigned to the empty set. Murphy provided an evidence average combination method. This method is based on the modified model and has faster convergence. The deficiency of this method is that it uses the unfused evidence simply, without considering the interconnectedness between them. Reference [12] used a modified Euclidean distance to determine the correlation between the evidence and to obtain the weight of evidence before modifying the original evidence and finally making the fusion decisions. Reference [13] pointed out that the use of a conflict coefficient to measure the conflict between the evidence was not sufficient: the conflict between the evidence is also related to the pignistic probability distance, so these two factors should be used together to measure conflict degree between the evidence. Reference [7] presented a conflicting interval evidence fusion method based on the evidence similarity measure. This method defines the probability conversion rules of the extended pignistic probability and converts the interval evidence into interval pignistic probability and uses the fuzzy interval to normalize the Euclidean distance and gets the similarity between interval pignistic, thus determining the similarity matrix between any two evidences

and obtaining confidence of the interval evidence. Different situations of evidence conflict are classified and discussed according to the size of the two values. However, Reference [13] did not consider the distance of the evidence body, so the conflict representation model is still not comprehensive. More importantly, there is no analysis of the conflict and the relationship leading to the conflict and it has not been developed to give a further approach to the process conflict.

Given the above problems, we propose an evidence synthesis method based on dimensionless indexes combined with KNN to improve the reliability and the rationality of the results. This paper is organized as follows. Sections 2, 3, and 4 introduce the problem statement, theory, and rotating machinery fault diagnosis experiment, respectively. Finally, a conclusion is drawn in Section 5.

2. Problem Statement

Vibration monitoring signals from rotating machinery in the event of a fault include nonlinear, random, and nonergodic information, which leads to great difficulty in fault signal analysis. Although a dimensional index is sensitive to the fault characteristics, its value will increase with the development of the fault; when the working conditions (load, speed, etc.) change, it is easy to be affected, resulting in unstable performance. The dimensionless index is not sensitive to disturbance of the vibration monitoring signal, performance is stable. Of dimensionless indexes, the pulse index and the kurtosis index are more sensitive to impact type fault, especially in the early stage of fault, since there is less pulse and no significant increase in other parameter values while the kurtosis index and pulse index rise fastly. Therefore, these two indexes are more sensitive in the early stage of fault in rotating machinery, resulting in increased fault interval range so it is difficult to distinguish.

Definition 1 (see [2, 3]). A dimensionless index is made up of the ratio of two amounts with the same dimension. When describing a particular system, it has a certain physical meaning. In fault diagnosis for dimensionless parameter indexes,

$$\Delta r_x = \frac{\left[\int_{-\infty}^{+\infty} |x|^T \xi(x) dx \right]^{1/T}}{\left[\int_{-\infty}^{+\infty} |x|^m \xi(x) dx \right]^{1/m}}, \quad (1)$$

where x is the vibration amplitude and $\xi(x)$ is the probability density function of vibration amplitude. Multiple historical monitoring data of a single fault can be calculated using this function.

Our Aim. We attempted to answer the question regarding how to integrate reliable and reasonable results when the fault diagnosis of the rotating machinery fault signal is uncertain.

In this paper, we addressed the diagnoses of rotating machinery fault for large petrochemical enterprises. Sensors collected many kinds of fault data by mechanical operation in real time online, and the distance between this data and the known training samples was calculated using

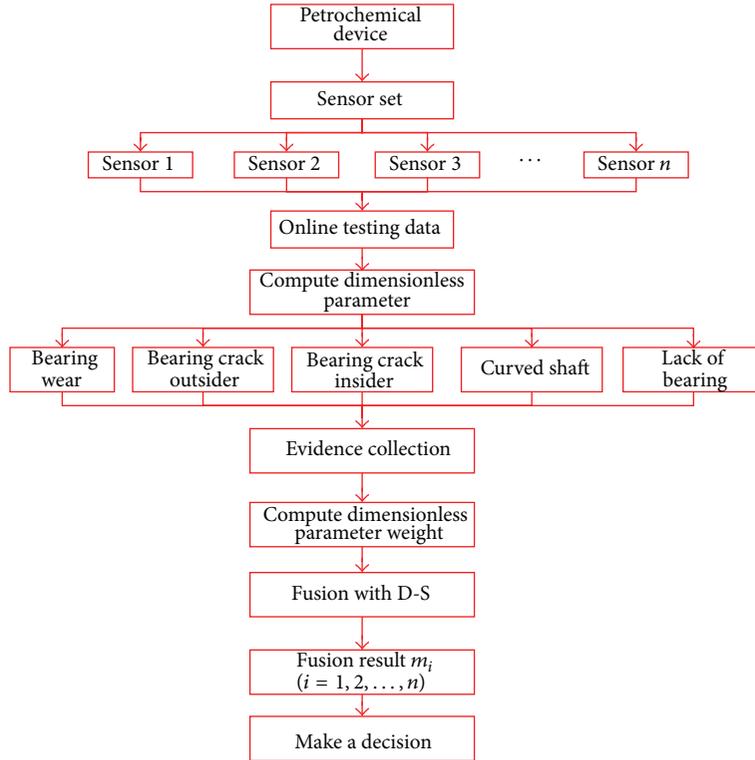


FIGURE 1: Flowchart of rotating machinery fault evidence synthesis diagnosis.

the KNN algorithm. After obtaining the distance between the test samples and the known training samples, we took the reciprocal value of the distance, as the probability that the tested sample is the kind of training sample. We fused evidence using the D-S evidence theory synthesis method to make a final decision about the fault. The specific flow is shown in Figure 1.

3. Theory

3.1. Calculation of Dimensionless Indexes and Determination of the Fault Zone. In this paper, we processed vibration monitoring signal using the method of dimensionless calculation [14].

Hypothesis 1 (see [1, 2, 5]). Under Definition 1, and $T = 1$, $m = 1$, then the waveform index

$$S_f = \frac{\left[\int_{-\infty}^{+\infty} |x|^2 \xi(x) dx \right]^{1/2}}{\left[\int_{-\infty}^{+\infty} |x| \xi(x) dx \right]} = \frac{\sqrt{E(|x|^2)}}{\sqrt{E(|x|)}}. \quad (2)$$

Similarly, (1) when $T = \infty$, $m = 1$, pulse index I_f is defined as

$$I_f = \lim_{T \rightarrow \infty} \frac{\left[\int_{-\infty}^{+\infty} |x|^T \xi(x) dx \right]^{1/T}}{\left[\int_{-\infty}^{+\infty} |x| \xi(x) dx \right]} = \lim_{T \rightarrow \infty} \frac{\sqrt[T]{E(|x|^T)}}{\sqrt{E(|x|)}}; \quad (3)$$

(2) when $T = \infty$, $m = 1/2$, margin index CL_f is defined as

$$CL_f = \lim_{T \rightarrow \infty} \frac{\left[\int_{-\infty}^{+\infty} |x|^T \xi(x) dx \right]^{1/T}}{\left[\int_{-\infty}^{+\infty} |x|^{1/2} \xi(x) dx \right]^2} = \lim_{T \rightarrow \infty} \frac{\sqrt[T]{E(|x|^T)}}{\left[\sqrt{E(|x|)} \right]^2}; \quad (4)$$

(3) when $T = \infty$, $m = 2$, peak index C_f is defined as

$$C_f = \lim_{T \rightarrow \infty} \frac{\left[\int_{-\infty}^{+\infty} |x|^T \xi(x) dx \right]^{1/T}}{\left[\int_{-\infty}^{+\infty} |x|^2 \xi(x) dx \right]^{1/2}} = \lim_{T \rightarrow \infty} \frac{\sqrt[T]{E(|x|^T)}}{\sqrt{E(|x|^2)}}; \quad (5)$$

(4) when $T = \infty$, $m = 4$, kurtosis index K_v is defined as

$$K_v = \lim_{T \rightarrow \infty} \frac{\left[\int_{-\infty}^{+\infty} |x|^4 \xi(x) dx \right]}{\left[\int_{-\infty}^{+\infty} x^2 \xi(x) dx \right]^2} = \frac{E(|x|^4)}{\left[E(|x|^2) \right]^2}. \quad (6)$$

A dimensionless index is made up of the ratio of two amounts with the same dimension. In this paper, we monitored signals based on the probability density function of the monitoring signal. This dimensionless index is a ratio, which is not affected by the magnitude of the signal, and the correlations between the sensitivity of vibration detector, amplifier, and the magnification are not large, so the monitoring system without calibration can be used in the actual equipment fault diagnosis [1, 14].

To use the dimensionless index in the study of fault diagnosis, we began with petrochemical core units. We collected data online in real time and calculated the normal state of the rotation unit and many kinds of dimensionless index parameters when each fault happens. Then we calculated the maximum value and minimum value of each dimensionless index for each of core units in the normal state and all kinds of fault states.

Hypothesis 2. N monitoring data of vibration data ε were collected under the single fault and N is relatively large.

Conclusion 1. Under the condition of Definition 1, Hypothesis 1, and Hypothesis 2, the expectation of the dimensionless index can be approximate

$$\varepsilon^{-T} = E(|\varepsilon|^T) = \frac{1}{N} \sum_{i=1}^N |\varepsilon_i|^T. \quad (7)$$

So, dimensionless index $\Delta\eta_x$ approximate

$$\Delta\eta_x = \frac{\sqrt[T]{\varepsilon^{-T}}}{\sqrt[m]{\varepsilon^{-m}}} \quad (8)$$

if $T = \infty$, then $\sqrt[T]{\varepsilon^{-T}} \approx \max_{j=1,2,\dots,N} |\varepsilon_j|$.

Conclusion 2. Under the condition of Definition 1, Hypothesis 1, and Hypothesis 2, sets of vibration monitoring data of a single fault history $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{N_k}$, where the value range is $\varepsilon^{-T} \in [\alpha_T, \beta_T]$, of ε^{-T} can be calculated; then the dimensionless index fault interval is

$$\begin{aligned} \Delta\eta_x &= \frac{\left[\int_{-\infty}^{+\infty} |x|^T \xi(x) dx \right]^{1/T}}{\left[\int_{-\infty}^{+\infty} |x|^m \xi(x) dx \right]^{1/m}} = \frac{\sqrt[T]{\varepsilon^{-T}}}{\sqrt[m]{\varepsilon^{-m}}} \in [c_\varepsilon, d_\varepsilon] \\ &= \frac{\left[\sqrt[T]{\alpha_T}, \sqrt[T]{\beta_T} \right]}{\left[\sqrt[m]{\alpha_m}, \sqrt[m]{\beta_m} \right]} \in \left[\frac{\sqrt[T]{\alpha_T}}{\sqrt[m]{\beta_m}}, \frac{\sqrt[T]{\beta_T}}{\sqrt[m]{\alpha_m}} \right]. \end{aligned} \quad (9)$$

3.2. KNN Algorithm. Cover and Hart proposed the K -nearest neighbor algorithm (KNN) in 1968 [15]. The idea behind the algorithm is to calculate the distance between tested samples and known training samples based on a distance function. Select k -nearest sample values and choose an unknown sample according to the k -nearest sample values. This method is widely used in fault diagnosis, text classification, data mining, machine learning, pattern recognition and image processing, and other domains. This paper has N fault samples distributed to c classes, S_f, C_f, I_f, CL_f , and K_v . Each class has N_i samples, $i = 1, 2, \dots, c$. We found k -nearest neighbors in all fault samples. k_i represents the number of k -nearest neighbors distributed to c class. The k -nearest neighbor judgment function is

$$g_i(x) = k_i, \quad (i = 1, 2, \dots, c). \quad (10)$$

3.3. Evidence Theory

3.3.1. Classic D-S Algorithm. D-S evidence theory is an uncertainty reasoning method, also known as belief function theory. It is widely used in intelligent data processing, information fusion, expert systems, data mining, fault diagnosis, target identification, decision analysis, and other domains. This theory provides useful evidence combination rules to fuse and update evidence information in order to solve the problem of processing uncertain information.

Evidence synthesis is the core of the evidence theory. It fuses independent evidence information coming from different information sources in order to produce more reliable evidence information. However, D-S evidence synthesis is limited in different degrees in practical application, especially when evidence conflicts highly or fully. In these cases D-S evidence synthesis loses efficacy, and so researchers at home and abroad in the field have proposed many improvements from their different perspectives. At present, China's fault diagnosis technology is widely used in military, aerospace, chemicals, shipbuilding, and so forth. There are many theories and methods of fault diagnosis, and evidential reasoning has a great significance in fault research. It contains uncertainty information processing, the effective integration of information, determinations of the credibility of the fault indicators, formation, and decision-making. In this paper, we use the idea of evidence theory combined with the dimensionless index to solve such uncertainty problems. Through multifeature fusion recognition analysis, we improve the recognition performance and accuracy of fault diagnosis using effective, appropriate diagnostic methods and determine the root cause of failure quickly [9].

In a large crew equipment, we can install the sensors in different parts of large crew of the equipment to achieve equipment testing. The information from sensors provides all the fault information from each part that needs monitoring and forms a body of evidence. Different evidence bodies correspond to different credits functions. Through analyzing credit functions, we can obtain the corresponding credit and fuse each credit function using certain D-S combination principles to determine the fault eventually.

(1) *Basic Probability Assignment.* In the recognition framework Θ , the basic probability assignment (BPA) is a $2^\Theta \rightarrow [0, 1]$ function m , called the mass function. This satisfies

$$\begin{aligned} m(\emptyset) &= 0, \\ \sum_{A \subseteq \Theta} m(A) &= 1, \end{aligned} \quad (11)$$

where the $m(A)$ which makes $m(A) > 0$ is called a focal element for A .

(2) *Trust Function.* Trust function is also known as belief function. In recognition framework Θ , based on BPA, trust function definition of m is

$$\text{Bel}(B) = \sum_{B \subseteq A} m(B). \quad (12)$$

(3) *Likelihood Function.* Likelihood function is also known as plausibility function. In recognition framework Θ , based on BPA, likelihood function definition of m is

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \quad (13)$$

(4) *Confidence Zone.* In evidence theory, hypothetical A is in recognition framework Θ

$$m(\Theta) = \sum_{B \cap A = \Phi} m(B),$$

$$m(A) = \frac{\sum_{\cap A_i = A} \prod_{i=1}^m m_i(A_i)}{1 - \sum_{\cap A_i = \Phi} \prod_{i=1}^m m_i(A_i)} = \frac{\sum_{\cap A_i = A} \prod_{i=1}^m m_i(A_i)}{1 - K}. \quad (14)$$

This equation is a classic synthetic formula from D-S evidence theory where the size of the K value which represents the conflict between all the evidence is called the normalization factor. The role of $1 - K$ is not to assign the nonzero probability values to the empty set in the process of evidence synthesis [16].

In the classical D-S evidence theory synthesis formula, especially for the case of a completed conflict (i.e., $K = 1$), the results obtained from (13) above are usually not consistent with the actual situation and the formula loses efficacy. People began to modify this method on the basis of the original formula. There are two main ways in which it can be modified [8].

(1) *Based on Modification Rules [9, 17].* The key to improving synthesis results is how to manage conflict. The new synthesis rules need to efficiently determine how to allocate conflict, and this problem also contains two small problems: which subsets should the conflict be reassigned to and after determining the subset, in what proportion should the conflict be allocated.

(2) *Based on Modification Evidence Source Modification [17].* This presumes that the D-S synthesis rules for evidence theory are not themselves wrong. When the evidence conflicts highly, evidence should be pretreated first, and then the D-S evidence theory synthesis rules should be used. For those evidence sources in which conflicts are great and unreliable, we can use the discount factor and other methods [18] to process the evidence source without modifying the synthesis rule.

3.3.2. *Improved D-S Algorithm.* Since the classic D-S theory can not manage conflict effectively, when evidence conflicts highly, the results using the D-S evidence synthesis rule is different from the actual situation. Many people in China have proposed various modifications to D-S evidence theory. Ye et al. proposed an evidence combination method based on the weight coefficients and the confliction probability distribution [16]. After calculating the weighting coefficients for each piece of evidence, the following evidence combination is used for information fusion. The steps are as follows [16].

(1) Calculate the degree of confliction K_{ij} between evidence K_i in evidence set K and other evidence E_j ($i = 1, 2, \dots, i - 1, i + 1, \dots, n$) and form confliction vector of K_i [16]:

$$K_i = (k_{i1}, k_{i2}, \dots, k_{ii-1}, k_{ii}, \dots, k_{in}) \quad (15)$$

wherein

$$K_{ij} = \sum_{A_i \cap A_j = \Phi, A_i \in E_i, A_j \in E_j} m_i(A_i) m_j(A_j), \quad (16)$$

$(i = 1, 2, \dots, n).$

(2) Process confliction vector K_i ($i = 1, 2, \dots, n$) with normalization:

$$K_i^N = \frac{(k_{i1}, k_{i2}, \dots, k_{ii-1}, k_{ii}, \dots, k_{in})}{\sum_{j=1, j \neq i} k_{ij}} \quad (17)$$

$$= (K_{i1}^N, K_{i2}^N, \dots, K_{ii-1}^N, K_{ii}^N, \dots, K_{in}^N).$$

(3) After normalization, calculate the entropy of confliction vector K_i^N :

$$H_i = \sum_{j=1, j \neq i} k_i^N \ln k_i^N \quad (i = 1, 2, \dots, n). \quad (18)$$

(4) Get countdown of entropy H_i :

$$H_i^{-1} = \frac{1}{H_i}. \quad (19)$$

(5) Calculate the weight coefficient of evidence E_i :

$$w_i = \frac{H_i^1}{\sum_{j=1}^n H_j^{-1}}. \quad (20)$$

After calculating the weighting coefficients of each evidence, the following evidence combination is used for information fusion; steps are as follow [16].

(1) Allocate the probability value to the proposition in the framework according to the evidence provided by evidence source and establish the weight vector of the evidence source:

$$W = (w_1, w_2, \dots, w_n). \quad (21)$$

(2) Assume $W_{\max} = \max(w_1, w_2, \dots, w_n)$; relative weight vector is available: $W = (w_1, w_2, \dots, w_n)/W_{\max}$, then we can determine the “discount rate” of the basic probability assignment value of the evidence. Using the “discount rate” to adjust basic probability assignment value of all the proposition in each recognition framework according to the following method, the basic probability assignment value after being adjusted is [16]

$$m_i^*(A_k) = \alpha_i m_i(A_i) \quad (22)$$

wherein the discount rate is

$$a_i = \frac{w_i^j}{w_{\max}}. \quad (23)$$

TABLE I: Various faults intervals (acceleration).

Fault index	Bearing wear	Bearing crack outsider	Bearing crack insider	Curved shaft	Lack of bearing
Waveform index	[1.284, 1.447]	[1.19, 1.327]	[1.172, 1.327]	[1.172, 1.312]	[1.19, 1.494]
Peak index	[3.140, 5.959]	[1.886, 4.818]	[2.215, 4.818]	[2.026, 3.686]	[2.503, 6.105]
Pulse index	[4.068, 8.625]	[2.320, 6.044]	[2.664, 6.044]	[2.460, 4.590]	[3.098, 8.602]
Margin index	[4.935, 10.76]	[2.683, 7.346]	[3.044, 7.346]	[2.832, 5.647]	[3.647, 10.72]
Margin index	[3.027, 10.79]	[2.531, 4.051]	[2.167, 3.85]	[2.167, 3.85]	[2.531, 8.378]

(3) By substituting the probability value of all propositions after adjusting into [11] formula, we can get a new synthetic formula:

$$\begin{aligned}
 m(\Phi) &= 0, \\
 m(A) &= p(A) + k \cdot q(A), \quad A \neq \Phi, \\
 p(A) &= \sum_{A \in E_i \cap \bigcap_{i=1}^n A_i = A} m_1^* A(1) m_2^* A(2) \cdots m_n^* A(n), \\
 k &= \sum_{A \in E_i \cap \bigcap_{i=1}^n A_i = A} m_1^* A(1) m_2^* A(2) \cdots m_n^* A(n), \quad (24)
 \end{aligned}$$

$$q(A) = \frac{1}{n} \sum_{i=1}^n m_i^*(A),$$

$$m(\Phi) = 1 - \sum_{A \subset \Theta} m(A).$$

New synthetic formula fully considers the importance of the fusion evidence that comes from different data and makes synthetic results more realistic. Moreover, except the above improved D-S evidence combination formula in detail, there are many other methods, such as D-S evidence combination formula based on credibility proposed by Sun et al. [19], an effective evidence theory synthesis formula proposed by Li et al. [11]. This not only reduces the confliction effectively but also makes the results of synthetic realistic.

3.4. Improved D-S Algorithm Application in the Rotating Machinery Fault Diagnosis. We have realized the rotating machinery fault diagnosis for large petrochemical enterprises. Through sensors collect all kinds of fault data by mechanical operation in real time online and calculate the distance between this data and the known training samples using KNN algorithm. After obtaining the distance between the tested samples and the known training samples, take the reciprocal value of the distance as the test sample probability and training sample probability. The specific flow of fusion evidences using D-S evidence theory synthesis method and making a final decision is shown in Figure 1. The implementation steps are as follows.

Step 1. Fault data can be collected from petrochemical rotation real time.

Step 2. Based on the collected failure data, dimensionless indexes can be calculated and fault zone (the maximum and

minimum range in 10 indices) can be set up. Use (2), (3), (4), (5), and (6) to calculate the waveform indices, peak indicators, pulse index, margin index, and the kurtosis index range faults.

Step 3. According to the KNN algorithm, the number of nearest fault points k can be found and the distribution can be derived.

Step 4. Use the improved D-S algorithm in (Section 3.3.2) to calculate degree of conflict (k_{ij}) and then conflict vector (K_i) can be obtained.

Step 5. Normalize the conflict vector (K_i ($i = 1, 2, 3, \dots, n$)) and calculate the (K_i^N) using (16).

Step 6. Calculate the entropy H_i in conflict vector (K_i^N) after normalizing. Meanwhile, the weighing values w_i of E_i can be calculated based on (18).

Step 7. Use (23) to correct D-S fusion data.

Step 8. Make final decisions after correction.

4. Rotating Machinery Fault Diagnosis Experiment

This experiment was conducted on large rotating machinery fault diagnosis experiment platform in petrochemical equipment fault diagnosis key laboratories of Guangdong province. Real time data collection of many kinds of fault types at Guangdong University of Petrochemical Technology (GDUPT) are shown in Figure 2. There are five causes of bearing fault in petrochemical rotary sets. There are bearing wear, bearing outer crack, bearing inner crack, bent axle, and lack of bearing. The rotating machinery vibration acceleration signal in the process of operation was detected and calculated using a linear operation to get the waveform indicator S_f , peak metric C_f , pulse index I_s , clearance factor CL_f , and kurtosis value K_v for each kind of fault.

In order to make the experimental data more accurate, we have collected 1024 fault points for every kind of fault and used them as the training samples. Five indexes of the training samples were obtained by linear operations, and the minimum and maximum values of the five indexes were selected to confirm the range of the indicators, as shown in Table 1. In Table 1, it can be seen in the sensitivity of



FIGURE 2: Rotating machinery fault diagnosis real experiment condition. (a) The developed real test bed. (b) Fault diagnosis rotating machinery test bed. (c) Normal bearing. (d) Wearing ball bearing. (e) Outer ring crack bearing. (f) Inner ring crack bearing. (g) Bend shaft. (h) Lacking ball bearing.

various indicators that the waveform index, because its scope is very small, is the least sensitive. By contrast, the sensitive of the margin index to the jamming signal is much higher. In addition, under the same kind of dimensionless index, the overlap of five kinds of faults is significant; that is, they are highly conflicted. For example, for bearing outer cracks and bearing inner cracks, the dimensionless index values range for the five kinds of indexes is generally low.

Choosing a group of data randomly from all the real time acquired data, for example, we can choose a bearing crack value of 3950 and use all the collected 1024 external bearing crack data to produce an array S , and we can get $S(5) = 1$, $S(12) = 5$. Then the data value of 3950 was subjected to a linear operation and the fault data values were obtained and used in KNN arithmetic. First, take the middle values of the five dimensionless indexes as the central values of the scope and then calculate the distance from the fault value to each central value. Here we will get

25 groups of distance values. Then convert distance to a probability value using KNN algorithm. The way we choose is to directly take the reciprocals of those 25 groups of distance values and obtain their corresponding probability values. The guiding ideology is that when a test samples is closer to a training sample, it has a higher probability to share the same category of that training sample. In order to make it meet the basic probability equation (1), a probability value normalized processing was performed in each index, and the results are shown in Figure 3. Figure 3 lists various fault probability values under the five indexes. Each indicator provides fault probability values for five kinds of faults including bearing wear, bearing outer crack, bearing inside crack, bent axle, and lack of bearing. We named each indicator to be a basic probability distribution function, which is also called the evidence collection. Five sets of evidence were formed by KNN algorithm, and the information from the 5 groups of evidence collection was fused using D-S evidence theory.

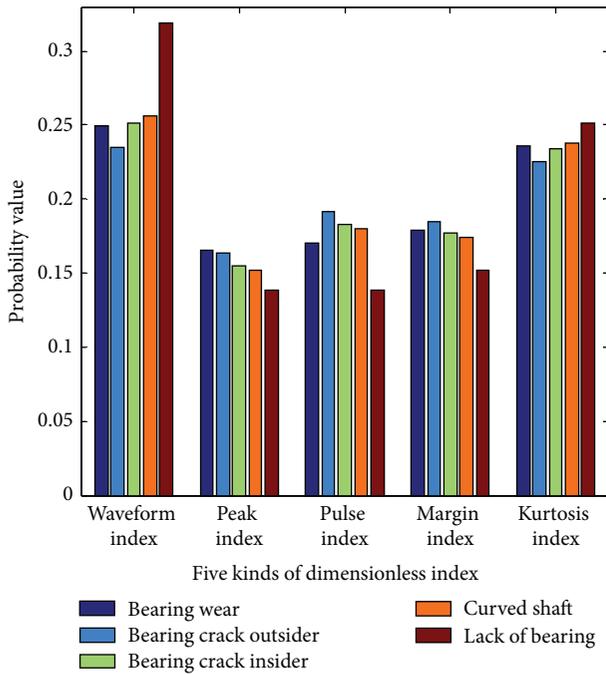


FIGURE 3: The results of KNN (The number of faults is 3950, $S(5) = 1$, $S(12) = 5$).

We used classic D-S evidence theory and various improvements to D-S evidence theory to match information fusion, and the results are shown in Figure 4. From Figure 4, we can see that the evidence collection processing is not strong enough when it meets the classic D-S evidence theory, especially the classic source of D-S evidence theory considers all of the evidences are equally important, it leads us to the even wrong conclusion with this situation [20]. In view of the above reasons, we used the improved D-S evidence theory, adding different weight coefficients to different evidence. The three methods in Figure 4 are based on the weight coefficient of the D-S evidence theory synthesis method. It can be seen that in comparing the three kinds of synthesis methods to the classical D-S evidence theory, that when evidence was highly conflicted, the other methods increase reliability and rationality of the results of synthesis. The tested data, however, were from an external bearing crack. Despite using improved D-S evidence theory the correct diagnosis of the fault was still not obtained.

We can see from Figure 4 that from the various sources of evidence, the probability value for the external bearing crack fault is not the largest. In other words, before fusing the evidence, each source of evidence does not think that it is the bearing outer crack that broke down, so the final fusion results are also incorrect.

5. Conclusion

There are some problems of identifying complex faults in petrochemical rotating machinery. First, the corresponding zone of the dimensionless index is difficult to determine.

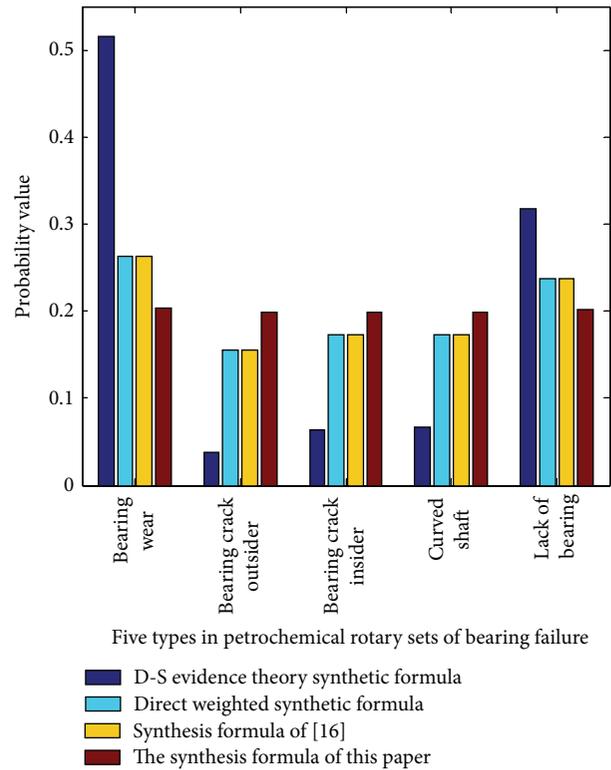


FIGURE 4: Evidence theoretical probability comparison table (the number of faults is 3950, $S(5) = 1$, $S(12) = 5$).

Second, when the data is transferred from the scene to a remote server, it is disturbed by various factors which cause transmission errors. Fluctuations in the calculation of the rotating machinery fault dimensionless indexes are large, resulting in difficulties with correct fault diagnosis. In this paper, we used a rotating machinery fault evidence synthesis diagnosis method combining dimensionless index with KNN to achieve fault evidence synthesis diagnosis of the rotating machinery to make the fusion result more reasonable and reliable. The increased reliability of the results will reduce the risk of decisions based on incorrect information.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the Associate Editor, Professor Gang Li, and Dr. X. S. Si giving them the opportunity to publish this correspondence paper. They would like to sincerely thank and acknowledge the wit-outputs and the tremendous work performed by the Associate Editor and the two anonymous reviewers for their insightful suggestions and thorough review, which greatly improve this correspondence paper. The authors are also grateful to Professor W. Xu, Professor L. Cai, Dr. H. Y. Wu, and Dr. Z. Zhang for their

help to improve writing quality. This work is supported by the National Natural Science Foundation of China under Grant nos. 61473331, 61271380, 61174113, and 61272382; the Natural Science Foundation of Guangdong Province of China (no. S2012010009870); the National Natural Science Foundation of Guangdong Petrochemical Equipment Fault Diagnosis Key Laboratory under Grant no. 643513; and the Guangdong University of Petrochemical Technology's Internal Project nos. 204341, 314004.

References

- [1] Q. H. Zhang, *Fault Diagnosis in Unit Based on Artificial Immune Detectors System*, China Petrochemical Press, 2008.
- [2] A. S. Qing, Q. H. Zhang, T. Y. Li, and Q. Hu, "The application of a compound dimensionless parameter for fault classifying of rotating machinery," *Modern Manufacturing Engineering*, no. 4, pp. 10–14, 2013.
- [3] Q. H. Zhang and Y. Z. Fu, "Research of adaptive immune network intrusion detection model," *International Journal of Systems, Control and Communications*, vol. 3, no. 3, pp. 280–286, 2011.
- [4] X.-S. Si, C.-H. Hu, J.-B. Yang, and Q. Zhang, "On the dynamic evidential reasoning algorithm for fault prediction," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5061–5080, 2011.
- [5] X. S. Si, C. H. Hu, and Z. J. Zhou, "Fault prediction model based on evidential reasoning approach," *Science in China, Series F: Information Sciences*, vol. 53, no. 10, pp. 2032–2046, 2010.
- [6] L. Zhang, J.-W. Liu, R.-C. Wang, and H.-Y. Wang, "Trust evaluation model based on improved D-S evidence theory," *Journal on Communications*, vol. 34, no. 7, pp. 167–173, 2013.
- [7] H.-S. Feng, X.-B. Xu, and C.-L. Wen, "A new fusion method of conflicting interval evidence based on the similarity measure of evidence," *Journal of Electronics and Information Technology*, vol. 34, no. 4, pp. 851–857, 2012.
- [8] H.-W. Guo, W.-K. Shi, Q.-K. Liu, and Y. Deng, "New combination rule of evidence," *Journal of Shanghai Jiaotong University*, vol. 40, no. 11, pp. 1895–1902, 2006.
- [9] R. R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93–137, 1987.
- [10] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Computational Intelligence*, vol. 4, no. 3, pp. 244–264, 1988.
- [11] B. C. Li, B. Wang, J. Wei, C. B. Qian, and Y. Q. Huang, "Efficient combination rule of evidence theory," *Journal of Data Acquisition and Processing*, vol. 17, no. 1, pp. 33–36, 2002.
- [12] D. Yong, S. WenKang, Z. ZhenFu, and L. Qi, "Combining belief functions based on distance of evidence," *Decision Support Systems*, vol. 38, no. 3, pp. 489–493, 2004.
- [13] W. Liu, "Analyzing the degree of conflict among belief functions," *Artificial Intelligence*, vol. 170, no. 11, pp. 909–924, 2006.
- [14] J. B. Xiong, Q. H. Zhang, G. X. Sun, Z. P. Peng, and Q. Liang, "Fusion of the dimensionless parameters and filtering methods in rotating machinery fault diagnosis," *Journal of Networks*, vol. 9, no. 5, pp. 1201–1207, 2014.
- [15] Y. Wang, *Study on text categorization based on decision tree and K nearest neighbors [Ph.D. thesis]*, Tientsin University, 2006.
- [16] Q. Ye, X.-P. Wu, and Y.-X. Song, "Evidence combination method based on the weight coefficients and the confliction probability distribution," *Systems Engineering and Electronics*, vol. 28, no. 7, pp. 1014–1081, 2006.
- [17] E. Lefevre, O. Colot, and P. Vannoorenberghe, "Belief function combination and conflict management," *Information Fusion*, vol. 3, no. 2, pp. 149–162, 2002.
- [18] B. He and H. L. Hu, "Multi-level DS evidence combination strategy," *Computer Engineering and Applications*, vol. 10, pp. 87–90, 2004.
- [19] Q. Sun, X. Ye, and W. K. Gu, "A new combination rules of evidence theory," *Acta Electronica Sinica*, vol. 28, no. 8, pp. 117–119, 2000.
- [20] J. B. Xiong, *Intelligence data fusion and its applications in ship dynamic positioning*, Guangdong university of technology [Ph.D. thesis], Guangdong University of Technology, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

