The Study on Rotating Machinery Early Fault Diagnosis based on Principal Component Analysis and Fuzzy C-means Algorithm

Qiang Zhao
Hebei University of Engineering, Handan, China, 056038

Abstract—This paper proposed an approach of mechanical failure information extraction and recognition in the early fault state variables, combined with the principal component analysis algorithm with FCM algorithm. Principal component analysis algorithm to get the characteristic value of data sets carries enough about fault in the time domain. Then FCM algorithm is used to analysis model is established in the classification feature different fault state. The application results of this method to identify variables deviation fault rotor test bed are acceptable.

Index Terms—Rotating machine; early fault diagnosis; PCA; FCM

I. INTRODUCTION

To ensure the safe operation of the rotating machinery, normal condition monitoring and fault state recognition technology should be given more and more attention. Now reports in the literature, all kinds of fault diagnosis technology are normal and fault state knows, can't identify the different countries the same mistake [1]. In fact, there are a series of degradation before most states the equipment. The United States estimates that need to identify and degradation of the classification accuracy is higher than the requirements of the fault diagnosis. Usually only for fault diagnosis and fault recognition of between normal classification including different states of the same mistake [1]. In fact, there are a series of degradation before most states the equipment. The United States estimates that need to identify and degradation of the classification accuracy is higher than the requirements of the fault diagnosis. Usually only for fault diagnosis and fault recognition of between normal classification including different states of the same mistake [1].

The state of the rotor system is a progressive fuzzy process from normal to normal, in this process the symptoms also blurred. Clustering analysis method and the theory of fuzzy mathematics foundation can provide a novel way to solve this kind of fuzzy problem. The fuzzy clustering analysis is effective pattern recognition. In many algorithm, the fuzzy c-means (FCM) method, fuzzy center) clustering algorithm, and has a good theory and high efficiency, has been successfully used in many fields, such as air compressor fault diagnosis, engine fault diagnosis (4, 5). So FCM cluster can be used for rotor system fault diagnosis of degradation early fuzzy process.

Components is such as gears, bearings and axis vibration in the operating process of collecting these components include valuable information about the machine or components conditions. However, the vibration measurement is usually a mixture several rotating components from the vibration of the special when they also faults, this makes the characteristics. Extraction and fault diagnosis. As an effective tool to identify the source of the signals mixed in the collection, blind source separation (BSS) [1] has been proved to be a very promising technology for the separation of the rotation Mechanical fault signal. Jas PuMa with [2] artists (second order application blind identify) method to separate two vibration signal water pump. South Africa [3] Fast ICA (fast forward position the separation of independent component analysis) method the vibration characteristics of the fault diagnosis motor. Gelle et al [4]. Separate the two source signal from the mixture of two signals with the accelerometer and a microphone Cardoso [5]. Put forward the traditional jade south approximate diagonalization Eigen-Matrices) method and Yeredor [6] put forward methods based on the second feature functions in general BSS problem. Methods by far the most BSS such as FastICA, artists, jade are based on the second or higher order statistical data source signals, however, rotating machinery. The fault signal such as bearing fault signal impulse-like Natural, the distribution of these signals just and stability [7] the distribution of order statistics mite no more than two, so serious degradation of separation Performance stability will happen to allocate resources to use these traditional BSS method. Only limited literature dedicated to the pulse signal BSS period, such as the foundation of the method. In the covariance statistics [8], jade method based on standardization statistics [9], this method based on the minimum standards is pension [10], usually used in speech signal separation. In this paper, we propose a novel BSS method. The separation of rotating machinery based on the fault signal content (the fractional lower order statistics), validation through the application of the
bearing separation properties the separation of the fault signal.

Fuzzy partition method for clustering c-partition to a single object, which is the number of predefined c cluster of [1]. One of the most famous fuzzy clustering method is divided into the fuzzy c-means algorithm (FCM) proposed by Bezdek j. c. in 1981 [2], this is a very powerful processing the ordinary data and uncertainty in real life meet [3, 4].

The fuzzy c-means algorithm is a kind of iterative algorithm, in order to achieve the required amount of the cluster C and initial clustering seeds must be predetermined. The seeds of the modified are every stage of the algorithm and each object in a certain degree of each cluster members to estimate. Each object is need and the seeds of cluster through the use of distance measure consider degree of members of the objects each cluster. The last of the algorithm, all given object is estimated that cluster degrees. K members. Atanassov [5, 6] the extended fuzzy sets, the concept of intuitionist fuzzy filter (IFS) in 1986, its outstanding characteristic is that it for each element in the IFS membership degree and as membership degree. The sum of its membership, as membership degree is not more than 1, if the sum equal to 1 so it is just a fuzzy sets, in other words, a intuition fuzzy sets a extended fuzzy sets. In addition, k. Atanassov and g. Gargov put forward the interval valued intuition fuzzy sets the concept (IVIFSs) in 1989 [7], the extension of IVIFS IFS.

Z. xu and j. wu is proposed based on intuition fuzzy c-means algorithm clustering algorithm and a interval valued intuition fuzzy c-means algorithm clustering algorithm of intuitionistic fuzzy data and interval valued intuition fuzzy data, said, IVIFCM IFCM algorithm respectively in 2010 [8]. However, and IVIFCM IFCM Algorithm can only be used for intuition fuzzy sets and interval valued intuition fuzzy sets, respectively is can not be used in any two n d point, from the definition of the distance between the two intuitionistic fuzzy set or interval valued intuition fuzzy set is a different definition of two n d point in the distance between. Entry in the IFS between 0 and 1, number, but a single entry in the n d point may be between 0 and 1 not the Numbers. The fuzzy c-means (FCM) clustering process of nitrogen and phosphorus space vector of input data, and use them, in conjunction with a necessary condition FCM objective function order to minimize, estimates that two sets of garland. The fuzzy c-means clustering method allows a data belonging to two or more than two clusters. Therefore, on the edge of the mass can be in cluster to a lesser degree scoring cluster center. This method is frequently used in pattern recognition.

II. THE EARLY FAULT FEATURE EXTRACTION BASED ON PCA

Principal component analysis (PCA) is a mathematical process, it uses a orthogonal transform a set of relevant factors of observations of may form a set of values of an independent variable of main component said. The number of main components is of less than or equal to the number of original data. Ownership statement is defined in this world a principal component has a as high variance (that is, accounts for as much as possible of the mutation data), each a successful components of the highest variance is constraints, it is probably right Angle (both) former components. Principal component that is independent, only in the joint data set the normal distribution. PCA is relative to the original variables scale.

Define a data matrix \( x = (x_1, x_2, K, x_p) \), whose covariance matrix is \( \Sigma \), then there must be a linear combination: \( y = l'x \). Among them, \( y = (y_1, y_2, K, y_p) \) is also the vector of dimension p, and \( l = (l_1, l_2, K, l_p) \) is the transformation matrix.

Equivalently, we are seeking to find the matrix Y, where Y is the Karhunen–Loève transform (KLT) of matrix X:

\[
\text{var}(y_j) = l_j'y_j = \sum_{i=1,2,K,p} l_i(i=1,2,K,p) \quad (1)
\]

\[
\text{con}(y_i, y_j) = l_i'y_j = \sum_{i=1,2,K,p} l_i(i=1,2,K,p) \quad (2)
\]

If the above equations satisfy the following conditions:

1) \( y_i \) and \( y_j \) (\( i \neq j \)) are uncorrelated;

2) \( y_i \) is the maximum variance vector of all linear combinations. \( y_2 \) Is one of vectors that are uncorrelated with \( y_1 \), and \( y_2 \) is the maximum variance vector among these vectors;

3) The coefficients of \( y_i \) = \( l_1'y \) satisfy the following equation. \( l_1'l = l(i=1,2,K,p) \)

So \( y_1, y_2, K, y_p \) are respectively called the first principal component, the second principal component..., the p th principal component. The principal component analysis is the simplest true eigenvector-based in multivariate analysis. Usually, its operation can be thought of as data reveals the internal structure of a kind of best explained variance data. If a multivariate data visualization as a set of relatively high dimensional data in space (1 shaft/variable), using principal component analysis (PCA) can provide the user with a projected picture, a "shadow" of the object when from its (in some sense) the most effective point of view. It is through the use of only the first few main components to transform the data dimension reduced

III THE FAULT STATE RECOGNITION BASED ON FUZZY C-MEANS (FCM) CLUSTERING

A FCM algorithm for the Basic Concepts

The fuzzy partition clustering method gives a single c-partition of the objects, where c is the predefined number of clusters [1]. One of the most well known fuzzy partition clustering methods is the fuzzy c-means algorithm (FCM) proposed by J. C. Bezdek in 1981[2], which is very powerful in dealing with nontrivial data and
uncertainties encountered in real life [3,4].

The fuzzy C-means algorithm is an iterative algorithm in which the desired number of clusters C and the initial clustering seeds has to be pre-defined. The seeds are modified in each stage of the algorithm and for each object a degree of membership to each of the clusters is estimated. Each object is needed to be compared with the cluster seeds by using a distance measure that takes into account the degree of membership of the object to each cluster. In the end of the algorithm, all the given objects are clustered according to the estimated membership degrees. K. Atanassov [5, 6] extended the notion of fuzzy set to intuitionistic fuzzy set (IFS) in 1986, whose prominent characteristic is that it assigns to each element in IFS a membership degree and a non-membership degree. The sum of its membership degree and non-membership degree is not larger than 1, if the sum is equal to 1 then it is just a fuzzy set, in other words, an intuitionistic fuzzy set is an extension of a fuzzy set. Furthermore, K. Atanassov and G. Gargov proposed the notion of interval-valued intuitionistic fuzzy sets (IVIFSs) in 1989 [7], and IVIFS is an extension of IFS.

Z. Xu and J. Wu proposed an intuitionistic fuzzy C-means clustering algorithm and an interval-valued intuitionistic fuzzy C-means clustering algorithm for intuitionistic fuzzy data and interval-valued intuitionistic fuzzy data, denoted IFCM and IVIFCM algorithm, respectively, in 2010 [8]. However, IFCM and IVIFCM Algorithm can only be used for intuitionistic fuzzy sets and interval-valued intuitionistic fuzzy sets, respectively, which cannot be used for any two n-dimensional points, since the definition of distance between two intuitionistic fuzzy sets or interval-valued intuitionistic fuzzy sets is quite different from the definition of distance between any two n-dimensional points. The entry in an IFS is a number between 0 and 1, but the entry in a n-dimensional point may be a number not between 0 and 1. The fuzzy c-means (FCM) clustering in the process of nitrogen and phosphorus space vector of input data, and use them, in conjunction with a order necessary conditions to minimize FCM target function, the estimate for two sets of a wreath. The fuzzy c-means clustering allows a piece of data belonging to two or more than two clusters. Therefore, on the edge of clumps, may be in the cluster to a lesser degree score of cluster center. This method is frequently used in pattern recognition.

Define data partitions set of n objects \( X = \{x_1, x_2, K, x_n\} \) in \( R^d \) dimensional space into c (1<c<n) fuzzy clusters with \( V_i = (i = 1, 2, c) \) cluster centers. The fuzzy clustering of objects is described by a fuzzy matrix \( U \) with n rows and c columns in which n is the number of data objects and c is the number of clusters. \( u_{ik} \), the element in the i th row and k th column in U, indicates the degree of association or membership function of the i th object with the k th cluster. The characters of U are as follows:

(1) \( \sum_{i=1}^{c} u_{ik} = 1, \forall k = 1, 2, K, n \)

(2) \( u_{ik} \in [0,1] \)

(3) \( 0 < \sum_{i=1}^{c} u_{ik} < 1, \forall i = 1, 2, K, n \)

\( V_i = (i = 1, 2, c) \) is i th cluster center vectors. The objective function of FCM algorithm is to minimize the Eq. (3)

\[
J_m(U, V) = \sum_{i=1}^{n} \sum_{k=1}^{c} (u_{ik})^m \| x_i - V_k \|^2, (1 \leq m \leq \infty)
\]

In which, m (m>1) is a scalar termed the weighting exponent and controls the fuzziness of the resulting clusters and \( \| x_i - V_k \| \), i=1,2,…,c is the Euclidian distance from object \( x_i \) to the cluster \( V_k \). The \( V_i \), center of the i th cluster, is obtained using Eq. (4).

\[
V_i = \frac{1}{\sum_{k=1}^{c} (u_{ik})^m} \sum_{k=1}^{c} (u_{ik})^n x_i, i = 1, 2, K, c
\]

The FCM algorithm is iterative and can be stated as follows:

(1) Select m (m>1); initialize the membership function values \( u_{ik} \), \( i = 1, 2, K, n \); \( k = 1, 2, K, c \)

(2) Compute the cluster centers \( V_i = (i = 1, 2, c) \) according to Eq.(4).

(3) Compute Euclidian distance \( \| x_i - V_k \| \), i=1,2,…,n, k=1,2…,c.

(4) Update the membership function \( u_{ik} \) according to Eq. (5).

\[
u_{ik} = \frac{1}{\sum_{j=1}^{n} (\| x_i - V_j \|)^{m-1}}, i = 1, 2, K, c
\]

(5) If not converged, go to step 2.

Several stopping rules can be used. One is to terminate the algorithm when the relative change in the centered values becomes small or when the objective function, Eq. (3), cannot be minimized more. The FCM algorithm is sensitive to initial values and it is likely to fall into local optima.

B The Fault Degradation State Recognition based on PCA-FCM

The typical rotating machinery fault a series of degradation state from normal to fail. The same fault, different types of the fault degree represent different states, fuzzy relationship exists in different states. The rotor axis deviation process is depicted on a state transition figure. The state cannot be directly confirm degradation, but a series of observation produced degradation process can reflect different state of existence and characteristics.
Fault states recognition model based on PCA-FCM is shown in Fig.2, and there are three steps: PCA feature extraction, FCM clustering analysis, fault state recognition. 1) A single time-domain eigen value of the collected signal can not reflect minor changes in fault state, so we carry out PCA to obtain PCs that carry most information of the several time-domain eigen values and the PCs can reflect the changes of fault states. 2) Collect sample data of n degradation states of the sand get the corresponding clustering centers 3) by using fuzzy closeness degree to recognize the different fault states, calculate Euclidean distance between each sample and fault clustering centers in the paper, thus we can obtain the results of diagnosis.

Figure 1. The rotor shaft misalignment of the state transition process

Figure 2. Fault states recognition model based on PCA-FCM

IV THE CASE STUDY

A The Fault Feature Extraction by PCA

The experimental subject, ZT - 3 rotor vibration simulation test-bed (shown in fig.3), is to simulate the rotating machinery rotor system fault-Shaft misalignment. The different size feeler gauges are installed at one end of the shaft bearing to simulate different degrees of the shaft misalignment faults. Using the data collecting instrument-EDES 4, developed by China University of Petroleum (Beijing) Fault Diagnosis Laboratory and vibration sensor to collect the data, and the sampling frequency is 2 KHz. The rotor speed is controlled by speed governor of the test-bed. In every misalignment fault state (20 μm, 60 μm, 120 μm) 20 sets of data are collected with the test-bed in 110V voltage. 10 sets are for FCM analysis; the other 10 sets are for testing the method accuracy.

B The Fault States Recognition based on FCM

Typically, there are many the rotor vibration signal description indexes; we take eight commonly used time domain eigen values, \( \{X_{pp}, X, X_{rms}, \sigma^2, \alpha, s_f, CL_f, K_u\} \) as original feature set of rotor fault. Where, \( X_{pp} \) is peak-to-peak value; \( X \) is absolute mean; \( X_{rms} \) is root-mean-square value (RMS); \( \sigma^2 \) is variance; \( \alpha \) is skewness; \( s_f \) is shape index; \( CL_f \) is clearance index; \( K_u \) is kurtosis index; and the features of signal are extracted by PCA. Though the 10 groups of training sample set for PCA, we obtain the first principal component and the second principal component, whose cumulative contribution rate are 98.55%; So the PCs carry the information to reflect the changes of e eight time domain eigen values.

© 2013 ACADEMY PUBLISHER
And then the results of this step are PCA-FCM models for different states of shaft in the rotator system. The rest ten data sets are going to be used for testing the trained model. The largest number of iterations step for training is settled at 100, and the convergence error of algorithm is 0.00001. The relationship of iterations step and the values gained from objective function is shown in figure 5. According to figure 5, the algorithm meets the standard of error settled while the data sets were training, and the number of iterations step is nine. Therefore, the main advantages of this method come from its high clustering speed and high clustering precision. The results acknowledged from the principle component clustering of three different fault conditions are shown in Fig.6.

![Image](image_url)

**Figure 5.** The convergence speed curve of the FCM algorithm

![Image](image_url)

**Figure 6.** The results of clustering with FCM algorithm

On the left of the data set (test data sets) are input to the training model, the distance between Europe, these test sets clustering center of the first three solve can calculate the fault condition were shown in table 1. Each test set, three European distance (distance and 20-microns clustering center, distance and 60-microns clustering center, the distance between the 20-microns clustering center and 120-microns) are calculated respectively. And the smallest distance, this means that the test set belong to which fault state. Have 30 test data set (every 10 said a fault state) are classified. Fault classification of states get rotor

<table>
<thead>
<tr>
<th>Sample Numbers</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Recognition accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-microns</td>
<td>0.2051</td>
<td>0.0711</td>
<td>0.4858</td>
<td>0.6572</td>
<td>0.9952</td>
<td>0.2786</td>
<td>0.4077</td>
<td>0.1392</td>
<td>0.3447</td>
<td>0.6276</td>
<td>100%</td>
</tr>
<tr>
<td>60-microns</td>
<td>7.2311</td>
<td>7.9244</td>
<td>5.1601</td>
<td>5.057</td>
<td>5.0772</td>
<td>9.4969</td>
<td>9.9833</td>
<td>6.4616</td>
<td>5.0804</td>
<td>3.9911</td>
<td>100%</td>
</tr>
<tr>
<td>120-microns</td>
<td>39.6566</td>
<td>43.2683</td>
<td>39.4614</td>
<td>39.4784</td>
<td>39.9403</td>
<td>46.8592</td>
<td>46.7772</td>
<td>41.8454</td>
<td>34.372</td>
<td>31.377</td>
<td>100%</td>
</tr>
</tbody>
</table>

**TABLE I.**

**THE CLASSIFICATION RESULTS OF TEST DATA SETS**

<table>
<thead>
<tr>
<th>Sample Numbers</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>Recognition accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-microns</td>
<td>5.9904</td>
<td>8.1845</td>
<td>8.7556</td>
<td>6.307</td>
<td>8.5582</td>
<td>8.8111</td>
<td>6.27</td>
<td>5.8836</td>
<td>3.3531</td>
<td>3.1033</td>
<td>100%</td>
</tr>
<tr>
<td>60-microns</td>
<td>5.144</td>
<td>0.1109</td>
<td>0.2362</td>
<td>0.7073</td>
<td>0.153</td>
<td>0.1898</td>
<td>0.0137</td>
<td>0.2009</td>
<td>0.3482</td>
<td>1.0048</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Numbers</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>Recognition accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-microns</td>
<td>41.4915</td>
<td>35.1741</td>
<td>38.7147</td>
<td>45.5687</td>
<td>39.9027</td>
<td>45.1911</td>
<td>22.2812</td>
<td>12.7537</td>
<td>39.4128</td>
<td>44.4483</td>
<td>100%</td>
</tr>
<tr>
<td>120-microns</td>
<td>2.7486</td>
<td>1.8118</td>
<td>1.0154</td>
<td>0.3738</td>
<td>0.6693</td>
<td>11.5581</td>
<td>0.967</td>
<td>0.2428</td>
<td>0.0462</td>
<td>0.24</td>
<td>100%</td>
</tr>
</tbody>
</table>

© 2013 ACADEMY PUBLISHER
experiment test is acceptable. To ensure the safe operation of the rotating machinery, the normal condition monitoring and fault state recognition technology should get more and more attention. Now reports in the literature, various kinds of fault diagnosis technology is normal and fault state know, can't identify the different countries the same mistakes [1]. In fact, there are a series of degradation of the several states before equipment. The United States estimates that need to identify and degradation of the classification accuracy is higher than the demand of the fault diagnosis. Usually only for fault diagnosis and fault recognition between different states of the normal classification including the same wrong, this is a problem of the fuzzy degradation process. Principal component analysis (PCA) is a typical statistical data analysis method. PCA is widely used in fault diagnosis, feature extraction, such as sensor fault diagnosis, statistical model evaluation method [2]. Therefore, PCA can be used to extract features of rotating machinery fault early. The state of the rotor system is a gradual process, from the normal normal fuzzy, in this process, the symptoms gradually fuzzy. Cluster analysis method and the theory of fuzzy mathematics foundation to provide a new method to solve this kind of fuzzy problem. The fuzzy clustering analysis is effective pattern recognition. In many algorithm, the fuzzy c-means (FCM) method, fuzzy center) clustering algorithm, and has a good theory and high efficiency, has been successfully used in many fields, such as air compressor fault diagnosis, engine fault diagnosis (4, 5). So FCM cluster can be used for rotor system fault diagnosis of degradation process early fuzzy.

Components is such as gears, bearings, axial vibration in the operation to collect these components including information of value component machine or conditions. However, the vibration is usually a mixture of several rotating components of the vibration of the special, their shortcomings, it makes features. Extraction and fault diagnosis. As an effective tools to identify the source of the signals mixed in the collection, blind source separation (BSS) [1] has been proved to be a very promising technology of the separation of rotating machinery fault signal. The puma and [2] artists (second order application blind identify) method to separate two water pump vibration signal. South Africa [3] fast independent component analysis (ICA) (fast into the position of independent component analysis separation) method vibration characteristics of fault diagnosis of the motor. Gelle et al [4]. Separate the two source signal of two signals with the mixture of the accelerator and the microphone cardoso [5]. Put forward the traditional jade south approximate diagonalization Eigen-Matrices) method and Yeredor [6] proposed method based on the second feature functions general BSS problem. Methods by far the most BSS such as FastICA, artists, jade are based on the second or higher order statistics data signals, however, rotating machinery. Such as bearing fault signal fault signal impulse-like a natural distribution, the signal is [7] and stability of order statistics distribution no more than two mites, then the serious degradation of the separation of the performance stability will happen to use these traditional allocate resources BSS method. Only limited literature BSS section dedicated to the pulse signal, such as basic methods. In the covariance statistics [8], the method of standardized statistics based on jade [9], this method based on the minimum standard is annuities [10], usually used in speech signal separation.

REFERENCES