We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Bearing Fault Diagnosis Using Information Fusion and Intelligent Algorithms

Jiangtao Huang

Additional information is available at the end of the chapter

1. Introduction

Rotating machinery is very common in industrial systems, and it plays an important role in industrial development and economic development. With the rapid advancement in industry, rotating machinery is becoming more and more complex and require constant attention. Although the reliability and robustness of rotating machinery also have been improving, some occasional failure events of components often lead to unexpected downtime while resulting in huge losses. And rolling element bearing is often at the heart of these rotating machinery which suffers from fault more frequently. These faults may cause the machine to break down and decrease its level of performance [6]. So, it is urgent to diagnose the incipient errors exactly in these bearings.

In traditional fault diagnosis, a single sensor is always used to get the operation conditions of several machine components. The collected signal involves many correlated features [33]. During operating process, the machine set can generate many kinds of signals. And those approaches based on the vibration signal analysis are advantageous because of their visual feature, easy measurability, high accuracy and reliability [34]. Fault diagnosis using raw vibration signals, a wide variety of techniques have been introduced in recent years. There are mainly including signal processing methods and intelligent systems application. Signal processing methods are traditional methods which are still common used, such as wavelet and wavelet packet methods [23–25], empirical mode decomposition [15, 35], time-frequency distributions [7], blind source separation [29]. While intelligent system approaches for fault diagnosis are including artificial neural networks (ANNs) [36], support vector machines (SVMs) [33], adaptive neuro-fuzzy inference system (ANFIS) [19] and fuzzy technique [28], etc.. These approaches are based on one data source or individual decision system, and many researchers have realized and shown that an individual decision system with a single data source can only acquire a limited classification capability which may not be enough for a particular application [22]. So, it is necessary to combine multiple decision systems to carry on failure diagnosis.

Multi-sensor information fusion is an emerging interdisciplinary beginning in the military field, and it has already been successfully applied in many different areas. In the field of industrial equipment fault diagnosis, multi-source information fusion technology application is still in its early stage. Multi-sensor information fusion is divided into three levels: sensor level, feature level and decision level. And multiple classifier ensemble approach belongs to decision level information fusion. In the recent years, the use of multiple classifiers has gained a lot of attention and researches have continuously showed the benefits of using multiple classifiers to solve complex problems [4]. In contrast, the feature-level fusion has not probably received the amount of attention it deserves [32].

By using information fusion theory, this chapter will introduce some bearing fault diagnosis approaches. And these methods can divide into two categories: fault diagnosis based on feature-level fusion [11] and fault diagnosis based on decision-level fusion [14]. In the proposed fusion methods for bearing fault diagnosis, some intelligent algorithms are used for feature dimension reduction or pattern recognition. The feature-level fusion approach for bearing fault diagnosis is using gene expression programming (GEP), while the decision-level fusion approach using multiple classifier ensemble method. And the decision-level fusion approach is based on the new bearing fault diagnosis method [12] which uses empirical mode decomposition (EMD) and fractal feature parameter classification.

2. Bearing fault diagnosis using fractal feature parameter classification

Faulty and normal machine conditions are always treated as classification problems based on learning pattern from empirical data modeling in complex mechanical processes and systems [31]. In this approach, a general framework for applying classification methods to fault diagnosis problems includes two steps: representative feature extraction and pattern classification. Feature extraction is a mapping process from the measured signal space to the feature space. Representative features which demonstrate the information of fault are extracted from the feature space. Pattern classification is the process of classifying the extracted features into different categories by geometric, statistic, neural or fuzzy classifiers. And recently, the development of artificial intelligence techniques has led to their application in fault diagnosis area. Meanwhile, artificial neural networks (ANNs) and support vector machines (SVMs) have been successfully applied to the intelligent fault diagnosis of mechanical equipment [27].

In practice, the classical approach is not always reliable when the extracted features are contaminated by noise. And most intelligent fault diagnosis approaches are complex, especially in solving multiple fault diagnosis problems. In this section, a novel, simple, fast and reliable intelligent method for solving multiple fault diagnosis problem will be proposed. And this approach is based on EMD and fractal feature parameter extraction.

2.1. Methodology

Fractal dimension is considered right from its invention [21] to be a good parameter to characterize time sequences of values of natural variables. And a simple, fast and accurate method for calculating the fractal dimension of data's time sequences was presented by Sy-Sang liaw and Feng-Yuan Chiu [20]. This method considers that a time sequence of $2^M + 1$ values is separated by a constant time interval which is well fitted by a fractal function $f(t)$ in

the period $[0, T]$. Then, calculating the fractal dimension D of $f(t)$ by using the known values of $f(t)$ at $t_j = jT/2^M$. To achieve this aim, Liaw and Chiu first defined $L_k(f)$, the piecewise linear interpolation of level $k(k = 0, 1, 2, ..., M)$, to $f(t)$ as the union of the line segments connecting the points $[t_j, f(t_j)]$ and $[t_{j+1}, f(t_{j+1})]$, where $t_j = jT/2^k$, $j = 0, 1, 2, ..., 2^k$ (see Figure 1). And then they checked out how poor the interpolation function $L_k(f)$ is relative to the next level of interpolation $L_{k+1}(f)$. The error of $L_k(f)$ is defined as the sum of the absolute value of the differences of $L_k(f)$ and $L_{k+1}(f)$ at all $t_j = jT/2^{k+1} \equiv j\varepsilon_k$:

$$\Delta_k \equiv \sum_{j=0}^{2^{k+1}} |L_{k+1}(f(t_j)) - L_k(f(t_j))| = \sum_{j=odd}^{2^{k+1}} |f(t_j) - \frac{f(t_j - \varepsilon_k) + f(t_j + \varepsilon_k)}{2}|, \ t_j = j\varepsilon_k \quad (1)$$

Liaw and Chiu [20] found that the value Δ_k is proportional to $(\varepsilon_k)^{1-D}$ when k is large enough.
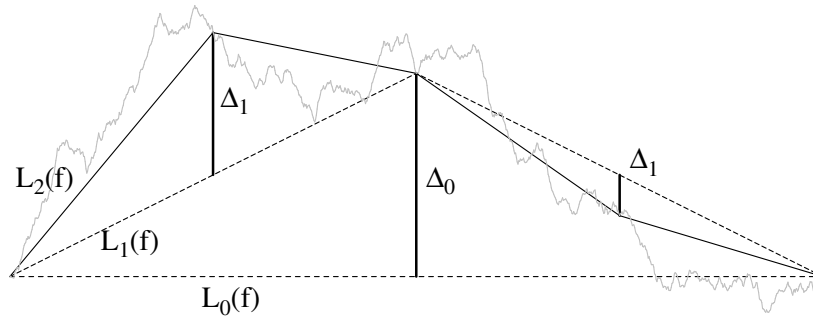
Figure 1. Piecewise interpolation $L_k(f)$ to a function $f(t)$ (grey) at level 0 (dotted), 1 (dashed), and 2 (solid). Δ_k (thick solid) denotes the error of the kth level interpolation with respect to the $k + 1$ level [20]

Thus, the fractal dimension D of $f(t)$ can be obtained from the slope s of the log-plot of Δ_k with respect to the level k by $D = 1 + s/log2$ for large enough k values.

In this bearing fault diagnosis method, raw vibration signal will be seen as a time sequences of data. Raw vibration signal is often heavily clouded by various noises due to the compounded effect of other machine elements' interferences and background noises presenting in the measuring device [2]. So, EMD is used to analysis raw vibration signal to filter noise before extracting its fractal feature. As discussed by Huang et al. [10], the EMD method is designated to deal with non-stationary and nonlinear signals. This method is based on the simple assumption that any data consists of different simple intrinsic modes of oscillations. Using the EMD method, complicated signals can be decomposed in a finite set of intrinsic mode functions (IMFs). Each IMF should meet the following two conditions: (1) in the whole data set of a signal, the number of extreme and the number of zero crossings must either equal or differ at most by one, and (2) at any time point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

Assume $x(t)$ is a vibration signal, and its empirical mode decomposition process can be described by following steps:

Step 1. Initialize: $r_0(t) = x(t), i = 1$.

Step 2. Extract the i-th intrinsic mode function (IMF) $c_i(t)$:

      Step 2.1. Initialize: $h_0(t) = r_{i-1}(t), j = 1$.

Step 2.2. Determine all the maximal values, minimal value points of $h_{j-1}(t)$ and fit all extreme points into the upper and lower envelope of the original signal with the cubic spline line.

Step 2.3. Determine the mean value of the upper and lower envelope of $h_{j-1}(t)$, designated as $m_{j-1}(t)$.

Step 2.4. Calculate the difference between $h_{j-1}(t)$ and $m_{j-1}(t)$, $h_j(t)$: $h_j(t) = h_{j-1}(t) - m_{j-1}(t)$.

Step 2.5. If $h_j(t)$ satisfies the conditions of IMF, then it is designated as $c_i(t) = h_j(t)$. Otherwise, update the value of j: $j = j + 1$, and return to Step 2.2.

Step 3. Get the remaining signal: $r_i(t) = r_{i-1}(t) - c_i(t)$, after decomposing the i-th IMF.

Step 4. When $c_i(t)$ or $r_i(t)$ satisfies the given termination condition, the cycle is ended. Designate the final remaining signal as $r_n(t)$ ($n = i$). Otherwise, update the value of i: $i = i + 1$, and return to Step 2.

Finally, raw vibration signal can be decomposed into n IMFs: $c_i(t), i = 1, ..., n$ and one residue function $r_n(t)$:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{2}$$

In this work, representative feature is fractal feature parameter extracting from each IMF. Because the method of fractal dimensions of time sequences needs k to be large enough, we use fractal feature parameter. And fractal feature parameter of each IMF will be calculated as Equation 3 shows. It is easy to know that the IMF's numbers of different raw vibration signal samples are different. And in the vibration signal examination, we find that the rich operating condition information is inside the front IMFs. So, we can integrate the residual IMFs into a component. In this new method, a parameter L is set to denote the number of IMF using to extract representative feature. And the L-th IMF will be re-denoted as $c_r(t)$ whose calculation form as Equation 4. Then, the feature set of each raw signal has L fractal feature parameters. For example, we set the value of parameter L as $L = 6$. Figure 2 summarizes all the IMFs and fractal features obtained from a bearing inner race fault signal sample. Table 1 presents the fractal feature parameters of IMFs of different operating condition vibration signal samples. And from Table 1, it is clear that fractal feature parameter sets of the same operating condition are similar, and it is easy to distinguish different operating conditions of fractal feature parameter.

$$p = \sum_{k=0}^{M-1} \Delta_k \tag{3}$$

$$c_r(t) = \sum_{i=L}^{n} c_i(t) \tag{4}$$

2.2. Results and discussion

By using fractal feature parameter classification, bearing fault diagnosis method is applied to the bearing fault signal analysis from the Case Western Reserve University website [3]. The ball bearings are installed in a motor driven mechanical system, as shown in Figure 3. By a self-aligning coupling, a three-phase induction motor is connected to a dynamometer
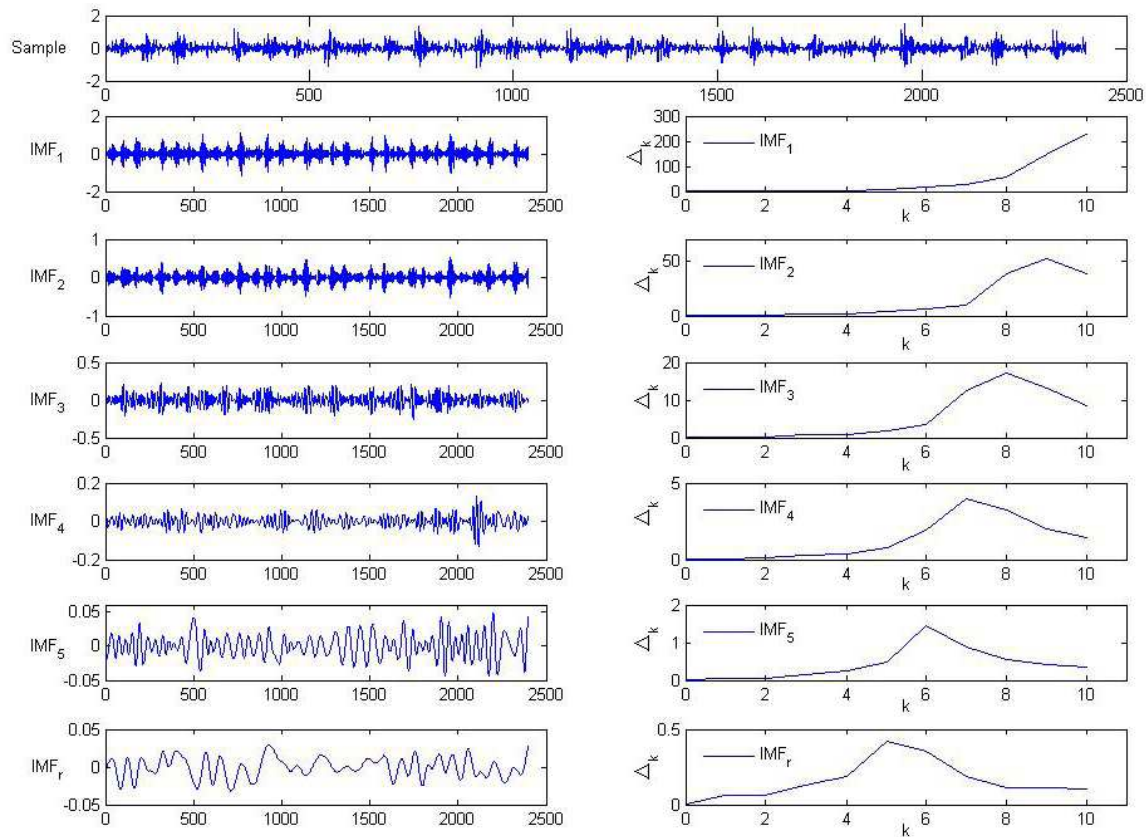
Figure 2. The resulting empirical mode decomposition components and fractal features from the inner race fault signal sample

Operating condition	Sample index	$p(c_1)$	$p(c_2)$	$p(c_3)$	$p(c_4)$	$p(c_5)$	$p(c_r)$
Normal	1	62.5909	26.8249	8.9951	6.6883	1.7284	0.6491
	2	67.8337	25.7203	8.9875	5.9469	2.8435	0.4777
Outer race	1	442.1013	45.3849	17.8250	5.9114	3.0012	0.6935
	2	479.2442	44.4325	19.1783	6.6492	3.4410	0.4973
Inner race	1	487.9925	151.8692	59.1485	14.1692	4.6696	1.5808
	2	511.5880	149.7561	68.9864	28.0217	5.8445	2.6485
Ball	1	295.3898	32.5473	19.6932	4.1315	2.0519	0.3888
	2	270.6188	33.8913	20.9217	4.4377	2.3469	0.3063

Table 1. Fractal feature parameters of different operating condition samples (defect size: 0.007inches)

and a torque sensor. The bearings are installed in a motor driven mechanical system. The dynamometer is under control so that desired torque load levels can be achieved. Vibration data is collected using accelerometer, which is attached to the housing with magnetic bases. Accelerometer is placed at the 12 o'clock position at the driven end of the motor housing. In machine condition monitoring, an accelerometer can provide rich information about conditions of several machine components. For example, the measured data from the accelerometer in this experiment is a mixture of signals reflecting conditions of the bearing inner race, outer race and rolling elements. The vibration data are collected by a 16 channel DAT recorder with 12,000 Hz.
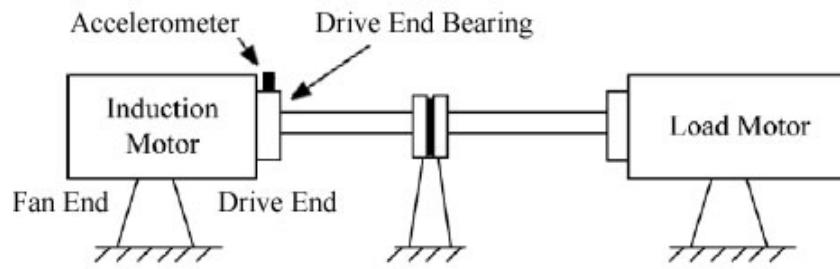
Figure 3. Schematic diagram of the experimental setup

As we referred in Figure 3, in the mechanical system, single point faults were introduced to the test bearings using electro-discharge machining with fault diameters of 7 mils, 14 mils and 21 mils. Each bearing was tested under four different loads (0, 1, 2 and 3 hp). Three bearing data sets (A-C) were obtained from the experimental system under the four different operating conditions: (1) under normal condition, (2) with outer race fault, (3) with inner race fault, (4) with ball fault. The detailed descriptions of the three data sets are shown in Table 2.

Data set	The number of training sample	The number of testing sample	Defect size(inches) (training/testing)	Operating condition	Class label
A	40	40	0/0	Normal	1
	40	40	0.007/0.007	Outer race fault	2
	40	40	0.007/0.007	Inner race fault	3
	40	40	0.007/0.007	Ball fault	4
B	40	40	0/0	Normal	1
	40	40	0.007/0.021	Outer race fault	2
	40	40	0.007/0.021	Inner race fault	3
	40	40	0.007/0.021	Ball fault	4
C	40	40	0/0	Normal	1
	40	40	0.007/0.007	Outer race fault	2
	40	40	0.007/0.007	Inner race fault	3
	40	40	0.007/0.007	Ball fault	4
	40	40	0.014/0.014	Outer race fault	5
	40	40	0.014/0.014	Inner race fault	6
	40	40	0.014/0.014	Ball fault	7
	40	40	0.021/0.021	Outer race fault	8
	40	40	0.021/0.021	Inner race fault	9
	40	40	0.021/0.021	Ball fault	10

Table 2. Description of three data sets

Data set A is formed by 320 samples. These samples include 4 different operating information under 4 conditions (0, 1, 2 and 3 hp), and among which the fault defect size is 0.007 inches. Every operating condition has 80 data samples. The whole data set is divided into 2 parts: 160 samples for training and 160 for testing. So, the task can be viewed as a four-class classification aimed at 4 different operating conditions. Data set B also contains 320 data samples. The training samples are including samples with 0.007 inches fault defect, while testing samples 0.021 inches fault defect. Data set B is used to further investigate the performance of fault diagnosis scheme. Data set C comprises 800 data samples including three different defect

sizes of 0.007, 0.014 and 0.021 inches under four different loads. It covers four different operating conditions, too. Each class data subset has been partitioned into two equal halves, one partition is used for training, while the other for testing. The purpose of data set C is to test the reliability of the novel approach in identifying the various grades of fault.

In order to evaluate the classification performance of the fractal feature parameter of IMF, orthogonal quadratic discriminant function (OQDF-E) [9] is used to train and test on three data sets showed in Table 2. Table 3 gives the classification performance on various data sets. The new bearing fault diagnosis method can get good decision accuracy as Table 3 shows. Table 4 extends the analysis of results and shows the classification performance between normal and fault operating condition. From Table 4, we can see that the new method using fractal feature parameter can get perfect performance in fault detection.

Data set	A	B	C
Train accuracy (%)	100	100	78.25
Test accuracy (%)	100	78.13	80

Table 3. Classification performance

	Data set A		Data set B		Data set C	
Operating condition	Normal	Fault	Normal	Fault	Normal	Fault
Test accuracy (%)	100	100	100	100	100	100

Table 4. Fault detection performance

3. Decision-level fusion for bearing fault diagnosis

In above section, we have proposed a simple, fast and good performance fault diagnosis approach. This approach is based on single sensor source and using individual classifier. It can obtain high accuracy on the multiple fault types recognition problems under the same fault degree. But when under multiple fault degrees, it declines in performance. To deal with this problem, this section will introduce a new method based on decision-level fusion for bearing fault diagnosis. The new fusion method includes four stages. These four stages are vibration signal acquisition and decomposition, fractal feature parameter extraction, single data source fault diagnosis and decision-level fusion for fault diagnosis. The first three stages are the same with the method described in the above section. So, we only state the last step in this section.

3.1. Methodology

Given a specific pattern recognition problem, different classifier has different classification performance. Very satisfactory results can not always be got if we simply conduct a study on a single classifier to improve its classification accuracy. Multiple classifier system (MCS) can overcome limitations of individual classifier and enhance classification accuracy. The techniques of combining the outputs of several classifiers have been applied to a wide range of real problems and it has been shown that MCSs outperform the traditional approach of using a single high-performance classifier [26].

The most often used classifiers combination approaches in MCS include the majority voting [30], the weighted combination (weighted averaging) [18], the probabilistic schemes [16, 17],

the Bayesian approach (naïve Bayes combination) [1, 18, 30], the Dempster-Shafer (D-S) theory of evidence [5, 30] and etc. This section will propose a new classifiers combination method which treats the combination process as linear programming problem.

Assume that K base classifiers are used in MCS, and M kinds of fault states including normal condition on the bearing fault diagnosis problem. Then, a decision matrix can be given as follow in the process of multiple classifiers combination.

$$D(x) = \begin{bmatrix} P_1(F_1|x) & P_1(F_2|x) & \dots & P_1(F_M|x) \\ P_2(F_1|x) & P_2(F_2|x) & \dots & P_2(F_M|x) \\ \vdots & \vdots & \ddots & \vdots \\ P_K(F_1|x) & P_K(F_2|x) & \dots & P_K(F_M|x) \end{bmatrix} \tag{5}$$

The new method introduced in this section will fuse those posterior probabilities in the decision matrix for constructing a global classifier E to make final decision. The posterior probability output of global classifier E for each fault state is calculated by following mode:

$$P_E(F_i|x) = \sum_{k=1}^{K} \beta_k P_k(F_i|x), \quad \forall i \in \{1, 2, \dots, M\}, \tag{6}$$

where β_k ($\sum_{k=1}^{K} \beta_k = 1$) is a dynamic association weight in MCS.

This new decision-level fusion method for bearing false diagnosis is based on the assumption: the base classifier has higher real-time recognition accuracy, if its posterior probabilities of all fault states are greater difference. That is to say, if individual decision system very determines that current operating condition belongs to a certain type of fault states, the posterior probability of the certain fault state will much higher than others. Using this hypothesis, the problem of multiple classifiers combination can be converted into a linear programming problem. And the objective function of this linear programming is defined as: $\sqrt{[\sum_{i=1}^{M}(P_E(F_i|x) - 1/M)^2]/M}$.

In current using classifier ensemble methods, base classifier's statistical performance is a major consideration factor. But we find the realtime decision information also can be a consideration factor. And in the new MCS method, we use within-class decision support [13] which is defined as: within-class decision support indicates that base classifier individual class recognition output gets the decision support degree from other same class recognition outputs in MCS. This decision support degree is measured by the difference between current output and its nearest output. For example, the within-class decision support of $P_k(F_i|x)$ which denotes posterior probability of the i-th state from the k-th base classifier is: $1 - \min_{1 \le k' \le K, k' \ne k} |P_k(F_i|x) - P_{k'}(F_i|x)|$.

Real-time decision support value (DSV) of base classifier in MCS is the sum of all class recognition output's within-class decision support value. And it is easy to get its calculation formula as:

$$\mu_k = \sum_{i=1}^{M} (1 - \min_{1 \le k' \le K, k' \ne k} |P_k(F_i|x) - P_{k'}(F_i|x)|), \quad \forall k \in \{1, 2, \dots, K\} \tag{7}$$

In the proposed decision-level fusion method for bearing fault diagnosis, we set a rule: if the real-time decision support value of base classifier is higher, its dynamic association weight of it is bigger. And this rule can be described as follows:

$$if \ \mu_k < \mu_{k'} \ then \ \beta_k < \beta_{k'}, \ \forall \, k, k' \in \{1, 2, \ldots, K\}, k' \neq k \tag{8}$$

$$if \ \mu_k = \mu_{k'} \ then \ \beta_k = \beta_{k'}, \ \forall \, k, k' \in \{1, 2, \ldots, K\}, k' \neq k \tag{9}$$

$$if \ \mu_k > \mu_{k'} \ then \ \beta_k > \beta_{k'}, \ \forall \, k, k' \in \{1, 2, \ldots, K\}, k' \neq k \tag{10}$$

That is to say, the relationship between dynamic association weights is determined by the relationship between real-time decision support values of different base classifiers. And these relationships will be used in the linear programming problem by the form of relationship vectors. Relationship vectors are defined as Table 5 shows (for example, $K = 3$). From Table 5, it is clear that each real-time relationship between DSVs is re-expressed by one or two relationship vectors. And it is also clear that each relationship vector is K dimensions. All these relationship vectors compose a relationship matrix which is denoted as R.

Real-time relationship between DSVs	Relationship vectors
$\mu_1 < \mu_2$	$[1 \ -1 \ 0]$
$\mu_1 = \mu_2$	$[-1 \ 1 \ 0] \ \& \ [1 \ -1 \ 0]$
$\mu_1 > \mu_2$	$[-1 \ 1 \ 0]$

Table 5. Example for relationship vector construction (K=3)

In order to simplify the fusion formula, a $K \times 1$ matrix β ($\beta = [\beta_1; \beta_2; \ldots; \beta_K]$) is used to replace β_k ($k \in \{1, 2, \ldots, K\}$) for calculating the decision output of global classifier E. Then, Equation 6 can be transformed into following simplified form: $P_E(x) = D(x)^T \beta$. And the objective function $\sqrt{[\sum_{i=1}^{M} (P_E(F_i|x) - 1/M)^2]/M}$ is simplified as: $\sqrt{\sum_{i=1}^{M} (P_E(F_i|x) - 1/M)^2}$, then further simplified to: $||D(x)^T \beta - \frac{1}{M}||_2$. Finally, use relationship matrix R to formulize constraint rules by the form: $R\beta \leq 0$. Now, we can give complete linear programming problem description as Equation 11 shows, where N is the count of relationship vectors of current relationship matrix.

$$max \quad ||D(x)^T \beta - \frac{1}{M}||_2$$

$$subject \ to \quad [1]_{1 \times K} \beta = 1$$

$$[0]_{K \times 1} \leq \beta \leq [1]_{K \times 1}$$

$$R\beta \leq [0]_{N \times 1} \tag{11}$$

Solving the linear programming problem as above, we can obtain the dynamic association weight matrix β. Using this dynamic association weight matrix, the fusion decision vector of global classifier E can be calculated. And the final decision of bearing fault diagnosis can be got by:

$$E(x) = i \ with \ P_E(F_i|x) = \max_{1 \leq i' \leq M} P_E(F_{i'}|x) \tag{12}$$

3.2. Results and discussion

The decision-level fusion method for bearing fault diagnosis is also applied to the rotating machinery from the Case Western Reserve University website [3]. In this experiment, vibration signals are collected from accelerometers which attached to the motor at different positions as Figure 4 shows. And dynamometer is used to control the torque load level. In this work, we study four different operating conditions recognition under four different loads (0, 1, 2 and 3 hp) with fault diameters of 7 mils, 14 mils and 21 mils. And these four operating conditions are normal condition, outer race fault, inner race fault and ball fault.
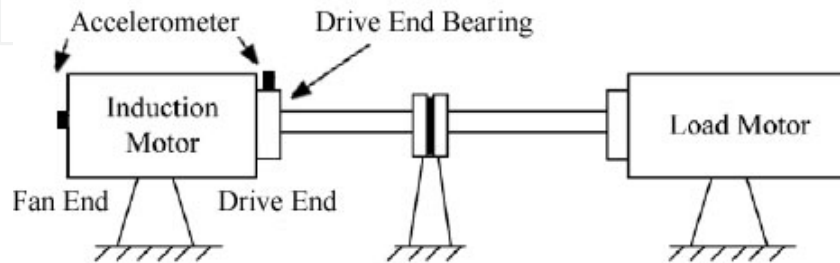
Figure 4. Schematic diagram of the experimental setup

Two data sets are constructed as Table 6 presents for testing the diagnosis performance of new decision-level fusion method. Each data set samples cover four different operating conditions and four different loads. And each class of two data sets has 160 data samples which are divided into two equal halves, one for training and the other for testing. Data set A is a four-class classification task corresponding to the four operating conditions. Data set B is a ten-class classification task corresponding to various grades of different faults.

Data set	The number of training sample	The number of testing sample	Defect size(inches) (training/testing)	Operating condition	Class label
A	80	80	0/0	Normal	1
	80	80	0.007/0.007	Outer race fault	2
	80	80	0.007/0.007	Inner race fault	3
	80	80	0.007/0.007	Ball fault	4
B	80	80	0/0	Normal	1
	80	80	0.007/0.007	Outer race fault	2
	80	80	0.007/0.007	Inner race fault	3
	80	80	0.007/0.007	Ball fault	4
	80	80	0.014/0.014	Outer race fault	5
	80	80	0.014/0.014	Inner race fault	6
	80	80	0.014/0.014	Ball fault	7
	80	80	0.021/0.021	Outer race fault	8
	80	80	0.021/0.021	Inner race fault	9
	80	80	0.021/0.021	Ball fault	10

Table 6. Description of two data sets

These data samples are extracted from two different sensor sources. And the number of samples from each sensor source is half of the total. If each sensor source's samples are seen as a subset of data set, each data set has two subsets. For example, data set A has two subsets:

A1 and A2. A1 is composed by the samples from the driven end accelerometer, while A2 from the fan end accelerometer. Table 7 gives the elements description of data set A in detail.

Data set	Class label	Sub dataset	Sensor source	The number of training sample	The number of testing sample
A	1	A1	driven end accelerometer	40	40
		A2	fan end accelerometer	40	40
	2	A1	driven end accelerometer	40	40
		A2	fan end accelerometer	40	40
	3	A1	driven end accelerometer	40	40
		A2	fan end accelerometer	40	40
	4	A1	driven end accelerometer	40	40
		A2	fan end accelerometer	40	40

Table 7. The elements description of data set A

In this work, two different classifiers, k-NN ($k = 7$) and Parzen classifier, are used for fault diagnosis task. And these two different classifiers identify rotating machinery operating condition using vibration signals collected from driven end and fan end accelerometers respectively. That is to say, each data set has four individual decision system results. And MCS is composed by these four base classifiers.

Table 8 gives individual classifier recognition accuracy on subsets of data set A and B. It is clear that individual classifiers can attain high bearing diagnosis accuracy on data set A, but they can not maintain the same high-performance on data set B whose fault diagnosis task is extended to various grades of different fault conditions.

Data set	Subset	Classifier	Training accuracy	Testing accuracy
A	A1	k-NN classifier	100%	100%
		Parzen classifier	100%	98.75%
	A2	k-NN classifier	100%	100%
		Parzen classifier	100%	99.38%
B	B1	k-NN classifier	88%	87.5%
		Parzen classifier	100%	85.75%
	B2	k-NN classifier	91.75%	88%
		Parzen classifier	100%	86.75%

Table 8. Fault diagnosis performance using base classifier

Table 9 shows the fault diagnosis performance of the novel decision-level fusion model using multiple classifier system. It is clear that the novel decision-level fusion model can get high recognition accuracy even in the difficult fault diagnosis task. In the testing phase of data set B, fault diagnosis accuracy of the new fusion model is higher than all base classifiers' accuracy as Table 8 shows. And it increases 6.5 percentage points averagely.

Data set	Training accuracy	Testing accuracy
A	100%	100%
B	100%	93.5%

Table 9. Fault diagnosis performance using the new fusion model

To further analyze performance of the new fusion model, a new k-NN classifier ($k = 3$) is added to multiple classifier system. The new multiple classifier system is used to test fault diagnosis performance on data set B. And sum rule is used to compare with the new approach. The comparison results are presented in Table 10. From Table 10, it is clear that the new approach attains the highest diagnosis accuracy.

7-NN classifier on B1	7-NN classifier on B2	Parzen classifier on B1	Parzen classifier on B2	3-NN classifier on B1	3-NN classifier on B2	Sum rule	New method
87.5%	88%	85.75%	86.75%	90%	87.75%	94.75%	95%

Table 10. Further comparison results of fault diagnosis performance

4. Feature-level fusion for bearing fault diagnosis

This section will propose a new multiple sources feature-level fusion model for bearing fault diagnosis using GEP. At present, the research of fault diagnosis based on feature-level fusion is still less, far from decision-level fusion attention. This is mainly because feature-level fusion is more difficult. But feature-level fusion application for fault diagnosis can be more effective to extract fault feature information. It is a way to improve the performance and robustness of bearing fault diagnosis system.

4.1. Methodology

GEP was invented by Ferreira [8], and it is the natural development of genetic algorithms and genetic programming. GEP uses linear chromosome which is composed of genes containing terminal and non-terminal symbols. Chromosomes can be modified by mutation, transposition, root-transposition, gene transposition, gene recombination, one-point and two-point recombination. GEP genes are composed of a head and a tail. The head contains function (non-terminal) and terminal symbols, while the tail contains only terminal symbols. For each problem, the head length (denoted h) is chosen by users, and then the head length is used to evaluate the tail length (denoted t) by: $t = (n - 1) \times h + 1$, where n is the number of arguments of the function with most arguments.

The flow of GEP is as follows:

Step 1. To set control parameters, select function classes, initialize population.

Step 2. To parse chromosome, evaluate population.

Step 3. To take use some operation such as selection, mutation, inserts sequence, recombine, mutation of random constant and inserts sequence of random constant to create new population.

Step 4. To implement best preservation strategy.

Step 5. If obtain most precision of computing, evolution would be finished, else turn to Step 2.

The new feature-level fusion model using GEP will be dealt with multiple sensors fusion problem. Assume that there are I sensors used in machine condition monitoring. For each sensor, the raw signal is divided into some signals by the same time segment. Each of these

signals is processed to extract some features. In this chapter, machine operating signal features only take into account the time-domain statistical characteristics. These feature parameters of time-domain are presented in Eequations. (13-23), where $x(t)$ is a signal series and N is its number of data points.

$$p_1 = \frac{1}{N} \sum_{n=1}^{N} x(n) \tag{13}$$

$$p_2 = \sqrt{\frac{\sum_{n=1}^{N} (x(n) - p_1)^2}{N - 1}} \tag{14}$$

$$p_3 = \sqrt{\frac{\sum_{n=1}^{N} x(n)^2}{N}} \tag{15}$$

$$p_4 = \left(\frac{\sum_{n=1}^{N} \sqrt{|x(n)|}}{N}\right)^2 \tag{16}$$

$$p_5 = \max |x(t)| \tag{17}$$

$$p_6 = \frac{\sum_{n=1}^{N} (x(n) - p_1)^3}{(N - 1) p_2^3} \tag{18}$$

$$p_7 = \frac{\sum_{n=1}^{N} (x(n) - p_1)^4}{(N - 1) p_2^4} \tag{19}$$

$$p_8 = \frac{p_5}{p_3} \tag{20}$$

$$p_9 = \frac{p_5}{p_4} \tag{21}$$

$$p_{10} = \frac{p_3}{\frac{1}{N} \sum_{n=1}^{N} |x(n)|} \tag{22}$$

$$p_{11} = \frac{p_5}{\frac{1}{N} \sum_{n=1}^{N} |x(n)|} \tag{23}$$

In the pattern recognition process of bearing fault diagnosis, we assume that there are M conditions including normal condition. Let S_m^i represents the set of all training samples belonging to m-th condition ($1 \leq m \leq M$) from the i-th sensor source. Feature-level fusion model is seek a way to fuse these features from different sensor sources. The new feature-level fusion model using GEP fuses these features by looking for a feature recognition function φ which maps the feature space to another space where samples in the same class are similarity and samples dissimilarity otherwise. And then, the feature recognition function φ will direct the building of a multi-source feature fusion model in reverse direction.

Functions $+$, $-$, $\times$, $/$, $sqrt$, exp are selected as input functions of GEP. The generation is set 5000, and fitness function is defined as:

$$Fitness = \frac{\sum_{m=1}^{M-1} \sum_{m'=m+1}^{M} (\sigma_m - \sigma_{m'})^2}{\sum_{m=1}^{M} \sum_{i=1}^{I} \sum_{k \in S_m^i} (\varphi(P_k^i) - \sigma_m)^2} \tag{24}$$

where σ_m is the mean of all m-th condition samples function mapping values, its formula is:

$$\sigma_m = \frac{1}{I} \sum_{i=1}^{I} \frac{\sum_{k \in S_m^i} \varphi(P_k^i)}{|S_m^i|} \tag{25}$$

After GEP training, a perfect feature recognition function φ can be got. Using function φ, we can calculate the mean mapping value of each operating condition samples from a certain sensor source. For building the multi-source feature evaluation matrix, the samples which are correctly classified are selected to calculate their mean. Multi-source feature evaluation matrix is composed by these mean values as Equation 26 shows.

$$
\begin{bmatrix}
\rho_1(1) & \rho_1(2) & \cdots & \rho_1(11) \\
\rho_2(1) & \rho_2(2) & \cdots & \rho_2(11) \\
\vdots & \vdots & \ddots & \vdots \\
\rho_M(1) & \rho_M(2) & \cdots & \rho_M(11)
\end{bmatrix}
\tag{26}
$$

In Equation 26, each element represents the mean value of each feature component of each operating condition. For example, $\rho_2(1)$ represents the mean of all correctly classified samples of the first feature from the 2-th operating condition.

4.2. Results and discussion

In order to evaluate the proposed feature-level fusion model, we apply it to bearing fault diagnosis. And data of this bearing fault diagnosis task are also take from a lab of the Case Western Reserve University website [3]. In this work, three experiments over three data sets are conducted as Table 11 shows. Those data are collected under various operating loads from motor driven end and fan end accelerometers.

Data set	The number of training sample	The number of testing sample	Defect size(inches) (training/testing)	Operating condition	Class label
A	80	80	0/0	Normal	1
	80	80	0.007/0.007	Outer race fault	2
	80	80	0.007/0.007	Inner race fault	3
	80	80	0.007/0.007	Ball fault	4
B	80	80	0/0	Normal	1
	80	80	0.007/0.021	Outer race fault	2
	80	80	0.007/0.021	Inner race fault	3
	80	80	0.007/0.021	Ball fault	4
C	80	80	0/0	Normal	1
	80	80	0.021/0.007	Outer race fault	2
	80	80	0.021/0.007	Inner race fault	3
	80	80	0.021/0.007	Ball fault	4

Table 11. Description of three data sets

Each data set covers four different operating conditions and four different loads (0, 1, 2 and 3 hp). And each class of data sets has 160 data samples which are divided into two equal

halves, one for training and the other for testing. The task of data set A is to identify different type of faults, while the experiment over data set B is carried out to further investigate the diagnosis performance of developing faults when the fusion model is trained by incipient faulty samples. And the experiment over data set C is to test the diagnosis performance of incipient faults when the fusion model is trained by the serious faulty samples.

Table 12 gives the results of these three experiments. From Table 12, we can see that the new feature-level fusion model using GEP can get stable, good diagnosis performance. And it is clear that testing performance is higher than training performance in the experiment on data set C. That is to say, when the new feature-level fusion model is trained by the serious faulty samples, it can easily identify incipient faults.

Data set	Training recognition accuracy	Testing recognition accuracy
A	83.75%	81.25%
B	83.75%	72.50%
C	76.25%	81.88%

Table 12. Fault diagnosis performance using feature-fusion model

In order to observe the performance change when the new feature-fusion model uses multiple source information instead of single source information, the new method is used to test bearing fault diagnosis performance with single sensor source. Table 13 gives the performance comparison result between more than one sensor (here using two sensors) and single sensor. From Table 13, we can see multi-sensor testing performance is greatly higher than the single sensor application using the new feature-level fusion model.

Data set	A	B	C
Multi-sensor testing performance increasing	0.56	0.48	0.57

Table 13. Performance comparison between multi-sensor and single sensor

5. Conclusion

This chapter has introduced some new methods for bearing fault diagnosis. These new approaches are using information fusion and intelligent algorithms. Bearing fault diagnosis is still an ongoing research subject over a decade and attracting a huge number of researchers in different areas. But most of those current using techniques mainly deal with single-source data. Many researches have shown that an individual decision system with a single data source can only acquire a limited classification capability which may not be enough for a particular application. So, we study a new way for bearing fault diagnosis using information fusion technology and intelligent algorithm.

Information fusion is a field still under research. Generally, information fusion process may happen in three levels: sensor level, feature level and decision level. Here, we propose a new feature level fusion method and a new decision level fusion method for bearing fault diagnosis. The feature level fusion method is using GEP which is a new intelligent algorithm. And it is a parallel fusion method. The decision level fusion approach is based on a new multiple classifier ensemble method. It analyzes raw vibration signal, and completes the feature extraction by using EMD and fractal feature parameter calculation. From experimental results, we can see that these new fusion model for bearing fault diagnosis task can get good

decision performance which is higher than the performance from traditional single sensor application.

Author details

Jiangtao Huang
College of Computer and Information Engineering, Guangxi Teachers Education University
Key Lab of Scientific Computing & Intelligent Information Processing in Universities of Guangxi,
China

6. References

[1] Altincay,H. (2005). On naïve Bayesian fusion of dependent classifiers. *Pattern Recognition Letters*, Vol. 26, No. 15, 2463-2473

[2] Bozchalooi, I.S.; Liang, M. (2008). A joint resonance frequency estimation and in-band noise reduction method for enhancing the detectability of bearing fault signals. *Mechanical Systems and Signal Processing*, Vol. 22, 915-933

[3] Case Western Reserve University, Bearing data centre. URL: http://www.eecs.cwru.edu/laborator/bearing

[4] Dara, R.A.; Kamel, M.S.; Wanas, N. (2009). Data dependency in multiple classifier systems. *Pattern Recognition*, Vol. 42, No. 7, 1260-1273

[5] Denoeux, T. (1995). A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics Part B*, Vol. 25, No. 5, 804-813

[6] Ericsson, S.; Grip, N.; Johansson, E.; Persson, L.E.; Sjömberg, J.O. (2005). Towards automatic detection of local bearing defects in rotating machines. *Mechanical Systems and Signal Processing*, Vol. 19, No.3, 509-535

[7] Fan, Y.S.; Zheng, G.T. (2007). Research of high-resolution vibration signal detection technique and application to mechanical fault diagnosis. *Mechanical Systems and Signal Processing*, Vol. 21, No. 2, 678-687

[8] Ferreira, C. (2001). Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems*, Vol. 13, 87-129

[9] Gu, S.C.; Tan, Y.; He, X.G. (2009). Orthogonal quadratic discriminant functions for face recognition. *Lecture Notes in Computer Science*, Vol. 5553, 466-475

[10] Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.L.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series anlysis. *Proceedings of the Royal Society of London Series*, Vol. 454, 903-995

[11] Huang, J.T.; Wang, M.H.; Xu, K.K. (2010). Rolling bearing fault diagnosis fusion model based on gene expression programming. *Journal of Information & Computational Science*, Vol. 7, No. 12, 2437-2442

[12] Huang, J.T.; Cao, X.W.; Li, W.J. (2010). Fault diagnosis of rotating machinery based on empirical mode decomposition and fractal feature parameter classification. *Key Engineering Materials*, Vols. 439-440, 658-663

[13] Huang, J.T.; Wang M.H. (2010). Multiple classifiers combination model for fault diagnosis using within-class decision support. *International Conference of Information Science and Management Engineering*, 226–229

[14] Huang, J.T.; Wang, M.H.; Li, W.J. (2011). A new decision-level fusion method for fault diagnosis. *International Journal of Digital Content Technology and its Applications*, Vol. 5, No. 3, 136-142

[15] Junsheng, C.; Dejie, Y.; Yu, Y. (2007). The application of energy operator demodulation approach based on EMD in machinery fault diagnosis. *Mechanical Systems and Signal Processing*, Vol. 21, No. 2, 668-677

[16] Kittler, J.; Hojjatoleslami, A.; Windeatt, T. (1997). Strategies for combining classifiers employing shared and distinct pattern representations. *Pattern Recognition Letters*, Vol. 18, No. 11-13, 1373-1377

[17] Kittler, J.; Hatef, M.; Duin, R.P.W.; Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, 226-239

[18] Kuncheva, L.I. (2007). *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience publication, New Jersey, USA.

[19] Lei, Y.; He, Z.; Zi, Y.; Hu, Q. (2007). Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. *Mechanical Systems and Signal Processing*, Vol. 21, No. 5, 2280-2294

[20] Liaw, S.S.; Chiu, F.Y. (2009). Fractal dimensions of time sequences. *Physica A*, Vol. 388, 3100-3106

[21] Mandelbrot, B.B. (1977). *Fractals, Form, Change and Dimension*, Freeman, San Francisco.

[22] Niu, G.; Han, T.; Yang, B.S.; Tan, A.C.C. (2007). Multi-agent decision fusion for motor fault diagnosis. *Mechanical Systems and Signal Processing*, Vol. 21, No. 3, 1285-1299

[23] Ocak, H.; Loparo, K.A.; Discenzo, F.M. (2007). Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics. *Journal of Sound and Vibration*, Vol. 302, No. 4-5, 951-961

[24] Purushotham, V.; Narayanan, S.; Prasad S.A.N. (2005). Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT&E International*, Vol. 38, No. 8, 654-664

[25] Rafiee, J.; Rafiee, M.A.; Tse, P.W. (2010). Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Systems with Applications*, Vol. 37, No. 6, 4568-4579

[26] Rasheed, S.; Stashuk, D.; Kamel, M. (2008). Fusion of multiple classifiers for motor unit potential sorting. *Biomedical Signal Processing and Control*, Vol. 3, No. 3, 229-243

[27] Samanta, B.; Al-Balushi, K.R. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical Systems and Signal Processing*, Vol. 17, 317-328

[28] Saravanan, N.; Cholairajan, S.; Ramachandran, K.I. (2009). Vibration-based fault diagnosis of spur bevel gear box using fuzzy technique. *Expert Systems with Applications*, Vol. 36, No. 2, 3119-3135

[29] Tse, P.W.; Zhang, J.Y.; Wang, X.J. (2006). Blind source separation and blind equalization algorithms for mechanical signal separation and identification. *Journal of Vibration and Control*, Vol. 12, No. 4, 395-423

[30] Xu, L.; Krzyzak, A.; Suen, C.Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on System, Man and Cybernetics*, Vol. 22, No. 3, 418-435

[31] Xu, Z.B.; Xuan, J.P.; Shi, T.L.; Wu, B.; Hu, Y.M. (2009). A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique. *Expert Systems with Applications*, Vol. 36, 11801-11807

[32] Yang, J.; Yang, J.Y.; Zhang, D. (2003). Feature fusion: parallel strategy vs. serial strategy. *Pattern Recognition*, Vol. 36, No. 6, 1369-1381

[33] Yang, J.Y.; Zhang, Y.Y.; Zhu, Y.S. (2007). Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mechanical Systems and Signal Processing*, Vol. 21, No. 5, 2012-2024

[34] Widodo, A.; Yang, B.S.; Han, T. (2007). Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert Systems with Applications*, Vol. 32, No. 2, 299-312

[35] Wu, F.; Qu, L. (2009). Diagnosis of subharmonic faults of large rotating machinery based on EMD. *Mechanical Systems and Signal Processing*, Vol. 23, No. 2, 467-475

[36] Wu, J.D.; Chan, J.J. (2009). Faulted gear identification of a rotating machinery based on wavelet transform and artificial neural network. *Expert Systems with Applications*, Vol. 36, No. 5, 8862-8875