Bearing Fault Diagnosis of Sorting Machine Induction Based on Improved Neural Network and Evidence Theory

Wei Chen, Qing-xuan Jia, Han-xu Sun, Si-cheng Nian
School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China
cschwbeijing@126.com

Abstract—Roller bearing is an important mechanical element of sorting machine induction. It usually has defects in outer race, inner race or balls due to continuous metal-metal contacts in high-speed operating conditions. This paper presents a novel diagnosis algorithm based on improved neural network and D-S evidence theory. Firstly, fault features are extracted through vibration signal analysis. Improved neural network classifier is then constructed to finish primary recognition, which introduces momentum to increase the learning rate. In order to reduce recognition uncertainty, each single classifier is regarded as independent evidence, and they are aggregated by improved Dempster’s combination rule. Experiment results show that proposed algorithm can improve diagnosis accuracy and decrease recognition uncertainty.

Keywords—fault diagnosis; roller bearing; neural network; evidence theory

I. INTRODUCTION

Roller bearing is a critical mechanical component and widely used in sorting machine induction. Because of high-speed operating conditions, it easily leads to bearing damage, such as bearing burns, metal falling of the surface, cage fragmentation [1]. The fault would affect normal operation, and seriously lead to a crash accident. It’s necessary to do condition monitoring and diagnose for roller bearing.

To realize reliable, fast and automated diagnostic procedure, various intelligent diagnosis techniques, such as artificial neural networks, acoustic emission, and support vector machine, fuzzy logic and evolving algorithms have been proposed [2-4]. Among many diagnosis methods, vibration signal analysis has been proven to be an effective approach for diagnosing roller bearing problem. The neural network, which has ability of strong nonlinear mapping, self-learning adapting, associative memory, information processing mode and excellent fault-tolerance performance, is extensively used in various applications, such as fault diagnosis, flood forecasting and estimate emission [5-7]. It can handle large dimensionality and non-linear characteristics problems effectively [8].

Dempster-Shafer (D-S) evidence theory has more rigorous reasoning process than probability theory. It provides an important way for expression and combination of uncertainty information, and has obtained widespread application in uncertainty reasoning, decision analysis and information fusion [9, 10]. It can be used as an effective method to dispose uncertainty diagnosis information from different measuring points. However, how to transform object and original data to basic probability assignment function objectively is a difficult problem. In order to overcome this, outputs of neural network can be used to construct basic probability in a unified identification framework. Through two level fusions, it may acquire novel information and higher recognition accuracy.

Based on above considerations, a novel diagnosis algorithm based on vibration signal analysis, neural network and evidence theory is proposed. The rest of the paper is organized as follows: Section 2 reviews the basic related theory. Section 3 introduces the novel algorithm in detail. Then numerical example is given to show the efficiency of the proposed approach in Section 4. Finally, a clear conclusion is drawn.

II. RELATED BASIC KNOWLEDGE

A. BP Neural Network

Neural networks are nets interconnected by lots of neurons, which use the nonlinear mapping transfer function. Many types of neural models have been proposed, such as feed-forward, feedback and self-organizing pattern neural network. Among them, multi-layer feed-forward BP neural networks (BPNN) is the ripest [11]. As the characteristic of BPNN, the precision and efficiency must be satisfied. It is usually composed by input layer, hidden layer and output layer. Its basic working principle is as following. Firstly, the network propagates input signals by weight to neurons in next layer. Then processes them with effectiveness function, transfers them from next connections to output units. If there are errors, then back-propagates and modifies weights by gradient descent method until errors meet demand. Finally, output information of neural network will be obtained.

The standard back-propagation method is based on the following popular gradient descent learning [12]:

\[
\Delta w_{ij}(n) = \alpha \delta_{j}(n)x_{i}(n)
\]

(1)

Where, \( \Delta w_{ij}(n) \) is the correction applied to weight \( w_{ij}(n) \), \( \alpha \) is learning rate parameter, \( \delta_{j}(n) \) is local gradient which points to the required changes in network weights, and \( x_{i}(n) \) is output of neuron \( j \) at iteration \( n \).

B. D-S Evidence Theory

In D-S evidence theory, all possible mutually exclusive and exhaustive propositions of the identifiable objects are enumerated in the frame of discernment \( \Theta \), expressed...
as $\Omega = \{\theta_1, \theta_2, \ldots, \theta_n\}$. The power set contains all $2^n$ possible subsets, noted by $P(\Omega) = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \ldots, \{\theta_n\}, \{\theta_1, \theta_2\}, \ldots, \{\theta_1, \theta_2, \ldots, \theta_n\}\}$, where $\emptyset$ denotes empty set.

A key point of evidence theory is the Basic Probability Assignment Function (BPAF). Each of evidence defines a mass function $m$, mapping the power set $P(\Omega)$ to the interval between 0 and 1, defined as $m : P(\Omega) \rightarrow [0, 1]$. The mass function $m$ satisfies following equations:

$$\sum_{A \in P(\Omega)} m(A) = 1, m(\emptyset) = 0$$  \hspace{1cm} (2)

Where $A$ is element of $P(\Omega)$, $m(A)$ represents the support degree of evidence. If $m(A) > 0$, $A$ is called focal element.

Belief function ($Bel$) and plausibility function ($Pl$) are defined respectively as follows:

$$Bel : P(\Omega) \rightarrow [0, 1], Bel(A) = \sum_{B \subseteq A} m(B)$$  \hspace{1cm} (3)

$$Pl : P(\Omega) \rightarrow [0, 1], Pl(A) = \sum_{B \supseteq A} m(B)$$

Because functions $m$, $Bel$ and $Pl$ have one-to-one correspondence, the probability of $A$ is $Bel(A)$.

D-S evidence theory also provides a useful combination rule between two evidences. Let $m_1$ and $m_2$ be two BPAFs in unified frame of discernment, their focus elements are $\{A_1, A_2, \ldots, A_j\}$ and $\{B_1, B_2, \ldots, B_j\}$ respectively. The classical D-S evidence combination rule is as follows

$$m(A) = m_1 \oplus m_2 \oplus \cdots \oplus m_i \cdots \oplus m_n(A)$$  \hspace{1cm} (4)

Where $\oplus$ called the orthogonal sum, the conflict degree $k = \sum_{A \cap B = \emptyset} m_1(A)m_2(B)$. If $k$ near to 1, it shows high conflict of evidence and may cause an illogical result. As for multiple evidences, the combination rule can be defined as:

$$m(A) = m_1 \oplus m_2 \oplus \cdots \oplus m_i \cdots \oplus m_n(A)$$  \hspace{1cm} (5)

### III. Diagnosis Algorithm Based Design

#### A. Structure of Algorithm

The algorithm’s structure is showed in Fig.1. As can be seen, it is made up of four parts: feature extraction, fault classification, evidence combination and decision making. The main reasoning process is as follows. Firstly, fault features are extracted from vibration signal at each channel. Then they are used as input eigenvector to train BPNN classifier. The evidence’s BPAF can be constructed according to BPNN’s output. All evidences are aggregated based on Dempster’s combination rule to make full use of multi-channel identifying information and resolve the contradiction between BPNN classifiers. Diagnosis results can be obtained according to max BPAF decision rule.

![Figure 1. Diagnosis model based on neural network and evidence theory](image)

#### B. Feature extraction

Feature selection has a significant impact on the result of pattern recognition. When roller bearing appears defect, impact vibrations are generated. It usually causes periodic impulses in vibration signals. Through analyzing the changes due to these impulses, it can recognize state of roller bearing. Amplitude and period of these impulses are determined by rotational speed, fault location, and bearing dimensions. Considering different fault locations as in Fig.2, the fundamental cage frequency $f_c$, ball frequency $f_{b_{axial}}$, inner race frequency $f_{i}$, and outer race frequency $f_{o}$ of these impulses can be obtained [13].

$$f_c = \frac{f}{2}(1 - \frac{d}{D}\cos(\gamma))$$

$$f_{b_{axial}} = \frac{D}{d}f_c(1 - \frac{d^2}{D^2}\cos^2(\gamma))$$

$$f_i = n_sf_{i} - f_c = \frac{n_s}{2}f_c(1 - \frac{d}{D}\cos(\gamma))$$

$$f_o = n_sf_{o} = \frac{n_s}{2}f_c(1 - \frac{d}{D}\cos(\gamma))$$  \hspace{1cm} (6)

Where $f_c$ is the shaft rotation frequency, $n_s$ is the number of rollers, $\gamma$ is a contact angle, $d$ and $D$ is the roller diameter and pitch diameter respectively as shown in Fig.2.
Kurtosis value and impulse factor are sensitive to impulse in vibration signal, they can be calculated as

\[
\text{Kurtosis} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^4 \left( 1 - \frac{1}{N} \sum_{i=1}^{N} x_i^2 \right)^2
\]

(7)

\[
\text{Im} f = \frac{1}{2N} \sum_{i=1}^{N} (\max(x_i) - \min(x_i))
\]

D. Evidence Combination and Decision

Each BPNN classifier at different measuring points can provide state assessment of roller bearing, combination is necessary to obtain more relevant information. According to D-S evidence theory, the frame of discernment in roller bearing diagnosis can be noted by \( \Theta = \{ A_1, A_2, A_3, A_4 \} \), where \( A_1, A_2, A_3 \) and \( A_4 \) denotes normal state, inner race, outer race and ball fault respectively.

As we know, D-S combination rule supposes all combined evidences have equal credibility. In fact, the reliability degree from different measuring points may be various. The signal to noise ratio (SNR) of raw vibration signal becomes worse with increasing distance between roller bearing and measuring points. Therefore, its credibility decreases. Based on above consideration, we define reasonable weights for BPNN classifiers using distance measure. Supposing there are \( N \) sensors around roller bearing. Let \( d_i \) \((i=1,2,...,N)\) be a distance between \( i^{th} \) BPNN and roller bearing. The credibility coefficient \( \alpha_i \) is as follows.

\[
\alpha_i = \min(d_i) / d_i
\]

(11)

Let \( O_i(j) \) \((i=1,2,...,N; j=1,2,...q)\) be \( j^{th} \) output value of \( i^{th} \) BPNN. Thus, the BPAF for the \( i^{th} \) evidence can be calculated as.

\[
m_l(A_i) = \frac{\alpha_i O_i(j)}{\sum_{j=1}^{q} O_i(j)}
\]

(12)

From the adjustments of probability assignment by new mass function, we can see that part of the support degree given to proposition \( A_j \) \(( A_j \neq \Theta ) \) is moved to \( \Theta \), which represents the degree of ignorance. In this way, the effect of less reliable sources on fusion result can be weakened. According to the above preprocessed BPAFs, evidences are aggregated based on evidence combination rule.

\[
\exists A_1, A_2 \in \Theta \text{, their basic probability assignment values are as follows:}
\]

\[
\begin{align*}
\{ m(A_j) = \max \{ m(A_j) \} \} \cup \{ m(A_j) = \max \{ m(A_j), A_j \neq A_k \} \}
\end{align*}
\]

(13)

If it satisfies \( m(A_j) - m(A_j) > \varepsilon \), where \( \varepsilon \) is a threshold, then \( A_j \) is decision result.
IV. EXPERIMENTAL ANALYSIS

This paper uses four types of roller bearings, no fault, ball fault, inner race fault and outer race fault. The relative parameters are shown in table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roller bearing</td>
<td>N205</td>
</tr>
<tr>
<td>Accelerometer</td>
<td>Lance ULT2052</td>
</tr>
<tr>
<td>Speed</td>
<td>1200 rpm</td>
</tr>
<tr>
<td>The number of rollers</td>
<td>12</td>
</tr>
<tr>
<td>Contact angle</td>
<td>0°</td>
</tr>
<tr>
<td>Roller diameter</td>
<td>0.75 cm</td>
</tr>
<tr>
<td>Pitch diameter</td>
<td>3.9 cm</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>25K</td>
</tr>
<tr>
<td>Length of sample</td>
<td>25K</td>
</tr>
<tr>
<td>Number of normal samples</td>
<td>200</td>
</tr>
<tr>
<td>Number of each fault samples</td>
<td>200</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.95</td>
</tr>
<tr>
<td>Objective exactitude</td>
<td>0.01</td>
</tr>
<tr>
<td>The number of input neurons</td>
<td>5</td>
</tr>
<tr>
<td>The number of output neurons</td>
<td>4</td>
</tr>
<tr>
<td>Training function</td>
<td>traindm</td>
</tr>
<tr>
<td>Activation function</td>
<td>Sigmod</td>
</tr>
<tr>
<td>Learning function</td>
<td>learmd</td>
</tr>
<tr>
<td>Threshold ε</td>
<td>0.1</td>
</tr>
</tbody>
</table>

According to (6), the defect frequency of outer race, inner race and ball is 96.92 Hz, 100.15 Hz and 143.07 Hz respectively. The number of hidden neurons belongs to interval [4, 13] with (8). Five fault features (kurtosis, impulse factor, ball defect frequency, inner race defect frequency, and outer race defect frequency) are used as input of BPNN classifier. Simulation results of training times and identification accuracy using various neurons in hidden layer is showed in table II. Therefore, structure is determined as 5–6–4 (input nodes, hidden neurons, and out nodes).

TABLE II. RESULTS UNDER VARIOUS HIDDEN NEURONS

<table>
<thead>
<tr>
<th>Number of neurons</th>
<th>Iteration times</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5.63 × 10⁴</td>
<td>7%</td>
</tr>
<tr>
<td>5</td>
<td>7.12 × 10⁴</td>
<td>6%</td>
</tr>
<tr>
<td>6</td>
<td>0.84 × 10⁴</td>
<td>5%</td>
</tr>
<tr>
<td>7</td>
<td>4.05 × 10⁴</td>
<td>10%</td>
</tr>
<tr>
<td>8</td>
<td>0.96 × 10⁴</td>
<td>9%</td>
</tr>
<tr>
<td>9</td>
<td>2.15 × 10⁵</td>
<td>13%</td>
</tr>
<tr>
<td>10</td>
<td>3.52 × 10⁵</td>
<td>8%</td>
</tr>
<tr>
<td>11</td>
<td>1.21 × 10⁵</td>
<td>6%</td>
</tr>
<tr>
<td>12</td>
<td>1.98 × 10⁵</td>
<td>7%</td>
</tr>
<tr>
<td>13</td>
<td>1.15 × 10⁵</td>
<td>9%</td>
</tr>
</tbody>
</table>

The distance between each sensor to roller bearing is 0.10m, 0.18m, 0.12m and 0.25m respectively. The corresponded classifier’s weight is 1.0000, 0.5556, 0.8333 and 0.4000. The results can be shown in table III with different BPNN classifier and evidence combination method.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPNN 1 classifier</td>
<td>95.75%(383/400)</td>
</tr>
<tr>
<td>BPNN 2 classifier</td>
<td>89.25%(357/400)</td>
</tr>
</tbody>
</table>

From table III, we can see that it improves the diagnosis accuracy of 99.5% (398/400) through evidence fusion and can avoid recognition uncertainty of single BPNN classifier.

V. CONCLUSION

A novel multi-fusion diagnosis algorithm for sorting machine induction based on improved BP neural network and D-S evidence theory is presented. Through simplifying structure of neural network, training rate and diagnosis accuracy were improved. Through taking each BPNN classifier as independent evidence and evidence fusion, it can solve problem of recognition indeterminacy of single BPNN classifier and improve diagnosis accuracy. Experiment results show that proposed fusion algorithm has high accuracy of 99.5%.

REFERENCES