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, multiplies output data of the connection by weight and 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]:

$$\Delta\omega_{ij}(n) = \alpha\delta_i(n)x_j(n) \tag{1}$$

Where, $\Delta\omega_{ij}(n)$ is the correction applied to weight $\omega_{ij}(n)$, a is learning rate parameter, δ_i is local gradient which points to the required changes in network weights, and $x_j(n)$ 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 Θ , expressed

as $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. The power set contains all 2^n possible subsets, noted by $P(\Theta) = \{\varnothing, \{\theta_1\}, \{\theta_2\}, \ldots, \{\theta_N\}, \{\theta_1, \theta_2\}, \{\theta_2, \theta_3\}, \ldots, \Theta\}$, where $\varnothing$ 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(Θ) to the interval between 0 and 1, defined as $m: P(\Theta) \to [0,1]$. The mass function *m* satisfies following equations:

$$\sum_{A \subset P(\Theta)} m(A) = 1, m(\varnothing) = 0 \tag{2}$$

Where *A* is element of $P(\Theta)$, $m(A)$ 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

$$\begin{cases} Bel: P(\Theta) \to [0,1], \ Bel(A) = \sum_{B \subseteq A} m(B) \\ Pl: \ P(\Theta) \to [0,1], \ Pl(A) = \sum_{B \cap A \neq \varnothing} m(B) \end{cases} \tag{3}$$

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

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

$$m(A) = m_1 \oplus m_2 = \begin{cases} 0 & A \neq \varnothing \\ (1-K)^{-1} \cdot \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j) & A \neq \varnothing \end{cases} \tag{4}$$

Where $\oplus$ called the orthogonal sum, the conflict degree $k = \sum_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j)$. 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:

$$m(A) = m_1 \oplus m_2 \oplus \cdots \oplus m_i \cdots \oplus m_N(A) \tag{5}$$

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.

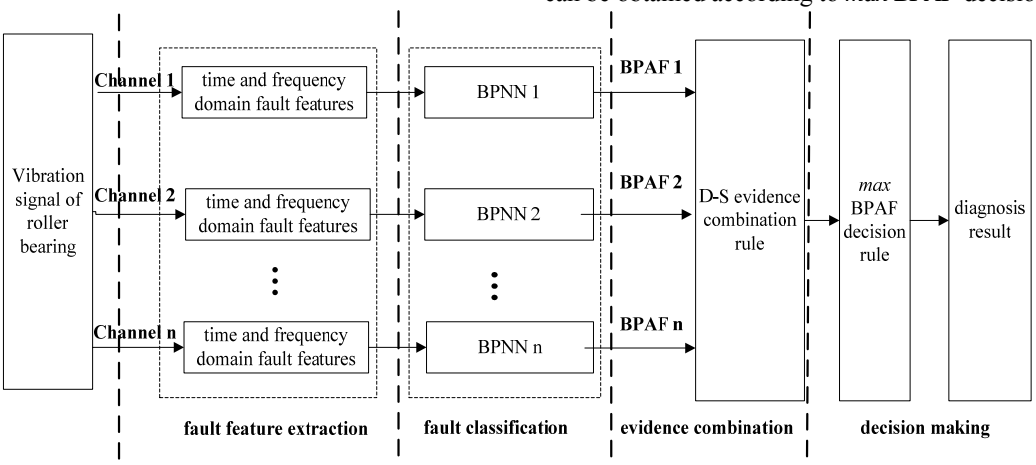

Figure 1. Diagnosis model based on neural network and evidence theory

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_{ball}, inner race frequency f_{ir}, and outer race frequency f_{or} of these impulses can be obtained [13].

$$\begin{cases} f_c = \frac{f_s}{2}(1 - \frac{d}{D}\cos(\gamma)) \\ f_{ball} = \frac{D}{d} f_s (1 - \frac{d^2}{D^2}\cos^2(\gamma)) \\ f_{ir} = n_b(f_s - f_c) = \frac{n_b f_s}{2}(1 - \frac{d}{D}\cos(\gamma)) \\ f_{or} = n_b f_c = \frac{n_b f_s}{2}(1 - \frac{d}{D}\cos(\gamma)) \end{cases} \tag{6}$$

Where f_s is the shaft rotation frequency, n_b is the number of rollers, γ 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

$$\begin{cases} Kurtosis = \dfrac{(1/N)\cdot\sum_{i=1}^{N}(x_i-\bar{x})^4}{\left(\sqrt{(1/N)\cdot\sum_{i=1}^{N}x_i^2}\right)^4} \\ \mathrm{Im}\, f = \dfrac{(1/2)(\max(x_i)-\min(x_i))}{(1/N)\sum_{i=1}^{N}|x_i|} \end{cases} \tag{7}$$

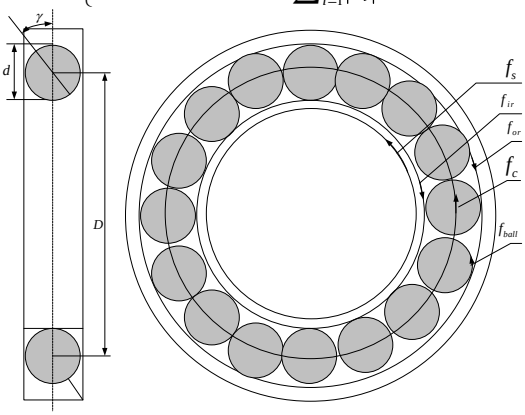

Figure 2. Bearing dimension and characteristic defect frequencies

Where x_i is amplitude at sampling point i, and N is the number of sampling points.

The above features are extracted from vibration signal and used as input eigenvector of BPNN classifier.

C. BPNN Classifier Design

BPNN's topology contains learning type, the number of hidden layers, learning rate and the number of neurons in hidden, input and output layers. Considering roller bearing has one normal state and three other fault states, such as ball defect, inner race defect, outer race defect, BPNN's output neurons should be 4. Its output objective vector denote by T = [1 0 0 0; 0 1 0 0; 0 0 1 0; 0 0 0 1]. Five fault features are used as input eigenvector of BPNN classifier. The activation function of the perceptrons is logsig in all cases.

Choosing reasonable neurons of hidden layer is important in neural network design, which is related to input neurons, output neurons and application. Optimization trial is applied to calculate in this paper. The range of neurons can be ensured according to following empirical formula.

$$n_1 = \sqrt{p+q} + \delta \tag{8}$$

Where parameter δ belongs to [1, 10], p, $n_1$and q are number of input, hidden layer and output neurons respectively. On the basis of same training exactitude, training and testing samples, optimal number of neurons in hidden is ensured according to training times and identification exactitude under different number of hidden layer neurons.

In addition, in order to speed up convergence, we would like to increase the learning rate. However, the algorithm will become unstable and oscillate back and forth when learning rate is rapid. According to the drawbacks of back propagation, an improved heuristic modification of back propagation is introduced, which adds momentum into change of parameter. The modified equations are as follows:

$$\Delta W^d(k) = \gamma\Delta W^d(k-1) - (1-\gamma)as^d(a^{d-1})^T \tag{9}$$

$$\Delta b^d(k) = \gamma\Delta b^d(k-1) - (1-\gamma)as^d \tag{10}$$

where γ is the momentum coefficient and set to 0.95, $W^d(k)$ 、 $b^d(k)$ and s^d is weight vector, bias vector, sensitivity function in the dth layer.

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 α_i is as follows.

$$\alpha_i = \min(d_i)/d_i \tag{11}$$

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

$$\begin{cases} m_i(A_j) = \dfrac{\alpha_i O_i(j)}{\sum_{j=1}^{q} O_i(j)} \\ m_i(\Theta) = 1-\alpha_i \end{cases} \tag{12}$$

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 Θ , 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.

$\exists A_1, A_2 \in \Theta$, their basic probability assignment values are as follows:

$$\begin{cases} m(A_1) = \max\{m(A_j), A_j \subset \Theta\} \\ m(A_2) = \max\{m(A_j), A_j \neq A_1\} \end{cases} \tag{13}$$

If it satisfies $m(A_1) - m(A_2) > \varepsilon$, where ε is a threshold, then A_1 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 I. PARAMETER SETTING

Parameter	Value
Roller bearing	N205
accelerometer	Lance ULT2052
speed	1200rpm
the number of rollers	12
contact angle	0°
roller diameter	0.75cm
pitch diameter	3.9cm
sampling frequency	25K
length of sample	25K
number of normal samples	200
number of each fault samples	200
learning rate	0.95
objective exactitude	0.01
the number of input neurons	5
the number of ouput neurons	4
training function	traindm
activation function	Sigmoid
learning function	learngdm
threshold ε	0.1

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

Number of neurons	Iteration times	Error (%)
4	5.63×10^4	7%
5	7.12×10^4	6%
6	$\mathbf{0.84\times10^4}$	**5%**
7	4.05×10^4	10%
8	0.96×10^4	9%
9	2.15×10^4	13%
10	3.52×10^4	8%
11	1.21×10^4	6%
12	1.98×10^4	7%
13	1.15×10^4	9%

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 III. RECOGNITION RESULTS WITH DIFFERENT METHODS

Methods	Accuracy (%)
BPNN 1 classifier	95.75%(383/400)
BPNN 2 classifier	89.25%(357/400)
BPNN 3 classifier	92.75%(371/400)
BPNN 4 classifier	81.25%(325/400)
Evidence combination (4 BPNNs)	**99.5%(398/400)**

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

[1] A.M. Al-Ghamdi, D.Mba. Comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. Mechanical systems and signal processing, 2006, 20(7):1537-1571.

[2] A.Choudhury, N.Tandon. Application of acoustic emission technique for the detection of defects in rolling element bearings. Tribology international, 2000, 33:39-45.

[3] S.Al-Dossary, R.I.R. Hamzah, D. Mba. Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing. Journal of Applied Acoustics, 2009, 70 (1):58-81.

[4] C. Liao, X. Li, D. Liu. Application of reassigned wavelet scalogram in feature extraction based on acoustic emission signal. Journal of Mechanical Engineering, 2009, 45 (2):273-279.

[5] Z.J. Li, Y. Zhou, and Z.K. Ma. River channel flood forecasting method of coupling wavelet neural network with autoregressive model [J]. Journal of Southeast University, 2008, 24(1): 90-94.

[6] D.C. Feng, J.M. Dias Pereira. Study on information fusion based on wavelet neural network and evidence theory in fault diagnosis. In: Proc of 8th International Conference on Electronic Measurement and Instruments, Xi'an, China: IEEE Press, 2007:3522-3526.

[7] E.O. Luis, M.T. Beatriz. A method to estimate emission rates from industrial stacks based on neural networks [J]. Chemosphere, 2004, 57(7): 691-696.

[8] T. Jin, P.W. Que, and T.L. Chen et al. Identification of magnetic flux leak inspection signals for pipeline detection using an improved BP neural network algorithm. Journal of Shanghai Jiaotong University, 2005, 39(7):1140-1144.

[9] A. Laha, N.R. Pal, and J. Das. Land cover classification using fuzzy rules and aggregation of contextual information through evidence theory [J]. IEEE Transaction on Geoscience and Remote Sensing, 2006, 44(6):1633-1641.

[10] H.W. Guo, W.K. Shi, Y. Deng et al. Evidential conflict and its 3D strategy: discard, discover and disassemble? [J]. Systems Engineering and Electronics, 2007, 29(6): 890 - 898.

[11] S.F. Jiang, C.M. Zhang, Z.W. Jin et al. Damage identification method based on back-propagation neural network and D-S evidence theory [J]. Journal of Shenyang Jianzhu University, 2007, 23(1):1-5.

[12] I. Aleksander, H. Morton. An introduction to neural computing [M]. 2nd edition, International Thomson Computer Press, London, 1995.

[13] N.Tandon, and A.Choudhury. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribology International, 1999, 32, 469-480.