A NOVEL APPROACH FOR INTEGRATED FAULT DIAGNOSIS BASED ON WAVELET PACKET TRANSFORM

Sheng Zhang[1], Joseph Mathew[2], Lin Ma and Yong Sun

School of Mechanical, Manufacturing and Medical Engineering, Queensland Univesity of Technology, Brisbane, QLD 4001, Australia

Abstract: Integrated machine fault diagnosis is usually conducted by considering different types of signals so as to improve the accuracy of diagnosis. This paper presents a novel approach for integrated machine fault diagnosis based on the vibration signals alone. Wavelet packet transform is adopted to analyze the vibration signals, followed by the selection of best bases. We consider each best basis as a local site, then extract features from it and make a local decision using probabilistic neural networks. The local decisions from each best basis are fused to be a global conclusion using a weighted average method. The whole diagnosis process is implemented under a uniform framework. An experimental case shows that this approach improves the accuracy of diagnosis.

1. INTRODUCTION

Wavelet transforms (WT) and wavelet packet transforms (WPT) are popular time-frequency analysis techniques [1-2]. In the past two decades, these techniques have been researched and applied in a variety of ways [3]. In vibration analysis, WT and WPT are preferred to the traditional fast Fourier transform (FFT) particularly in the analysis of transient signals [4-5].

WPT is the extension of WT and generates a binary tree of bases. Selecting the best basis from the tree is fundamental. For pattern classification, the best basis guarantees a best separation capability. In addition, extracting features from the best bases rather than from the binary tree helps reduce the feature dimensionality.

It is common to extract features from individual best basis, and then concatenate them in a high dimensional vector space. However, a high dimensional vector space may also be sliced into several low dimensional ones using distributed data mining (DDM) approach [6]. Decisions from each low dimensional space can be fused to a potentially more accurate conclusion. WPT creates opportunities for DDM and decision fusion, since it distributes the signal information into the best bases. In this paper the authors propose the extraction of features from individual best basis of WPT using the concepts of DDM. The local decisions are then made by classifiers. A final conclusion is drawn using the decision fusion technique. This approach was used to develop an integrated machine fault diagnosis procedure based on vibration signals.

The paper is arranged as follows. Section 2 describes the techniques used, viz., WPT, probabilistic neural networks, and decision fusion. Section 3 presents a framework for the integrated fault diagnosis. The proposed method is validated using signals acquired from typical faulty ball bearings in Section 4. A global probabilistic neural network using the combined features from all best bases is also adopted as a classifier for comparison. Section 5 contains the conclusions.

2. WPT, PROBABILISTIC NEURAL NETWORKS AND DECISION FUSION

2.1 Feature extraction from wavelet packet basis

WPT has a discrete format which is popularly used in engineering applications. To illustrate its underlying mathematical theory briefly, we denote $\{h_k\}_{k\in Z}$ and $\{g_k\}_{k\in Z}$ and as the quadrature mirror filter banks. A signal can be decomposed on the bases composed of functions of the form $2^{-j/2}u_n(2^{-j}t - k), j,k \in Z, n \in Z$ and

$$u_{2n}(t) = \sqrt{2}\sum_{n\in Z} h_k u_n(2t-k) \tag{1}$$

$$u_{2n+1}(t) = \sqrt{2}\sum_{n\in Z} g_k u_n(2t-k) \tag{2}$$

where j, k and n are the scale, time localization and oscillation parameters, respectively. $u_1(t)$ is the scaling function corresponding to a low-pass filter. The filtered signal is an approximation. $u_0(t)$ is the wavelet function corresponding to a high-pass filter. The filtered signal is a detail.

As the approximation and detail can be further sliced by dyadic decomposition, it can be seen that WPT generates a binary tree of bases. Each basis on the tree is indexed by a pair of integers (j, k). At the decomposition level j, there are $2^j - 1$ bases. The binary tree of bases can also be considered to form a 2-D time-frequency plane on which the signal information distributed. The information in the bases is redundant along two axes, i.e., information in child bases are overlapped with that in parent basis. It is preferable to select the best bases from the binary tree, so as to reduce the effort in data analysis without losing information.

[1] *Corresponding author, email: s2.zhang@qut.edu.au, fax: 61-7-38641469*

[2] *CEO, CRC for Integrated Engineering Asset Management*

The common best basis is usually used to identify signals which may come from different classes. For example, all signals are decomposed on their wavelet packet trees. A statistical measure of 'distance' is applied to produce a unique WPT-structured tree, from which the common best basis is identified [7-8]. For condition monitoring, characteristic wavelet packets can be selected based on statistical energy [9]. In current work, the unique WPT-structured tree was produced by the measure of cluster distance and the best basis was selected according to the Shannon entropy based criterion [10].

WPT creates opportunities for feature extraction and feature combination due to the rich information presented in the localized bases. Data mining, a convergence of knowledge discovering techniques [11], can play an important role in the extraction of features. Furthermore, the distributed best bases provide local sites for DDM. Based on the features from each best basis, local decisions can then be made by a classifier.

2.2 Probabilistic neural networks

Neural networks have been used successfully in pattern recognition as classifiers [12]. Popular neural networks include multilayer perception (MLP), radial basis networks (RBN), probabilistic neural networks (PNN), and self-organized maps (SOM). The PNN [13] is a special variant of RBN, which has found applications in solving regression and classification problems because it can be easily trained and can tackle applications with relatively few training samples.

A typical architecture of PNN is shown in Figure 1. It includes four layers. The first layer simply distributes the input to the pattern layer. In the pattern layer, usually each neuron corresponds to a training vector. The difference between the pattern x and the training vector is calculated in the neuron and then fed into a radial basis function, for which a Gaussian function is often used. Thus the output of neuron x_{ij} in the pattern layer is computed as

$$\phi_{ij}(x) = \frac{1}{(2\pi)^{d/2}\sigma^d}\exp[-\frac{(x-x_{ij})^T(x-x_{ij})}{2\sigma^2}],\ i=1,\cdots,m \quad (3)$$

where d denotes the dimension of the feature vector x, σ is the smoothing parameter, and m is the number of classes. The summation layer neurons calculate the maximum likelihood of pattern x and classify it into class C_i by summarizing and averaging the output of all neurons that belong to the same class

$$\overline{P}_i(x) = \frac{1}{(2\pi)^{d/2}\sigma^d}\frac{1}{N_i}\sum_{j=1}^{N_i}\exp[-\frac{(x-x_{ij})^T(x-x_{ij})}{2\sigma^2}] \quad (4)$$

where N_i denotes the total number of samples in class C_i. The probabilities given by Eq.(4) for each class are pooled in the output layer. This provides a way to assess the confidence that pattern x belongs to each class.

The PNN may include more neurons compared with MLP. For example, the pattern layer may include as many neurons as the number of training vectors. It may be noted that the PNN structure includes the smoothing parameter and the number of neurons, both of which can be optimized [14-15].

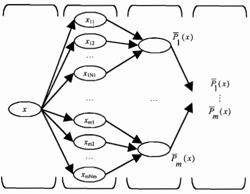

Figure 1. The architecture of a PNN

2.3 Decision fusion

Distributed data resources, such as distributed sensors, require the integration of local information to generate a final decision. The decision fusion technique improves the decision accuracy in pattern classification. The present work employs probabilistic neural networks for fault diagnosis. Local decisions are derived from each best basis of wavelet packets, which are then fused as a final decision at the classifier level [16]. Different methods are available for decision fusion, such as the weighted average method, winner-take-all principle, Bayesian rule, and Dempster-Shafer's method [17]. The weighted average method together with winner-take-all principle was adopted in this work.

3. PROCEDURE TO IMPLEMENT INTEGRATED FAULT DIAGNOSIS

The integrated fault diagnosis is based on vibration signal analysis using WPT for feature extraction. PNN is used for fault diagnosis on each best basis after which the local conclusions are fused. This procedure is implemented under a uniform framework as shown in Figure 2. The framework includes four parts in neural networks language: input layer, signal processing and feature extraction layer, PNN layer and decision fusion layer. Each part is explained as follows.

1) Signals are presented at the input layer.
2) The second layer is for signal processing and feature extraction. WPT is used to analyze the signals and n best bases are searched. The feature vector extracted from individual best basis is denoted as x. As mentioned above, each best basis is associated with a neural network for fault classification.
3) For each best basis, a PNN is employed to classify the feature vectors. The output of the ith PNN is a vector $P_i=[\overline{P}_1,\cdots,\overline{P}_m]^T$ whose elements given by Eq. (4) indicate how close the input is to each fault class.
4) The decision vectors from each PNN are combined to be a decision matrix $P=[P_1,\cdots,P_n]$ of size $m \times n$. If no strong

evidence shows that some best bases are more sensitive to the faults than the others, a weight vector in decision fusion layer can be set as

$$W = ones(m,1) \tag{5}$$

The decision fusion layer considers contributions from each PNN output and generates a fused probability $\overline{P}(C_i \mid x)$ representing the class the pattern x belongs to.

$$\overline{P}(C_i \mid x) = P * W,\ i = 1, \cdots, m \tag{6}$$

To make a final decision, we pick the maximum of the probabilities from $\overline{P}(C_i \mid x)$ and produce a 1 for that class and a 0 for the other classes - the winner-take-all principle.

$$P(C_i \mid \mathbf{x}) = \begin{cases} 1 & \max(\overline{P}(C_i \mid x)) \\ 0 & others \end{cases} \tag{7}$$

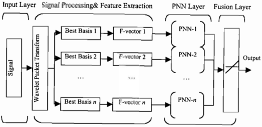

Figure 2. The integrated fault diagnosis framework

From the procedure, we note that all the necessary tasks are placed under the one framework. Since WPT and PNN are highly computational, they can be incorporated into an automatic integrated fault diagnosis procedure.

4. A CASE STUDY

Rolling element bearings are key components in mechanical systems. Their failures account for a large percentage of breakdowns in rotating machinery. Some of them can be catastrophic. Conducting diagnosis and prognosis on bearings is therefore fundamental to maintaining the integrity of mechanical systems.

Ready-made experimental data of rolling element bearing faults from Case Western Reserve University were used to test our methodology [18]. A single fault was introduced by electro-discharge machining on the outer-race, inner-race and ball, respectively. The collected data associated with the three types of faults came from different working conditions, i.e., under different RPM and loads. This ensured that the data are general in the sense that broad conditions are covered, which benefits the generalization of classifiers.

We adopted relatively few samples for testing our methodology. For example, in each fault class, 50 samples were used for classifier training, while 50 samples were used for classifier testing. Since three types of faults were involved, this resulted in 300 samples.

Following the procedure in Section 3, the signals were first decomposed by WPT up to level 3 by Db20 wavelets. Figure 3 illustrates a typical signal from a faulty outer-race and its WPT. Figure 4 shows six common best bases selected by the discriminate distance related Shannon entropy criterion for the three signal classes.

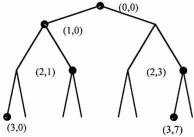

Figure 3. WPT for an outer race signal

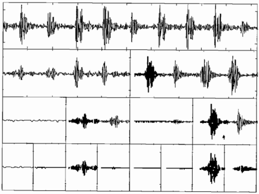

Figure 4. Common best basis

For a specific best basis, we selected signal energy, signal kurtosis and their combinations as the features respectively. The training and testing datasets consisted of 1-D or 2-D feature vectors. The PNN with the smoothing parameter $\sqrt{50}$ was used for each basis. The signals were then classified to make the local decisions, which were further fused to reach a final decision.

Table 1 provides the final classification results using the proposed approach. It is found that when signal energy is employed as the feature, all 50 testing signals in each class are correctly classified. However, when kurtosis is used as the feature, it leads to numerous misclassifications in each class. Using kurtosis and energy feature also deteriorate the classification results.

A feature vector can be constructed in that its elements come from different best bases. Instead of using the DDM approach, a global decision can be made based on this feature vector. A classifier is again required. We adopted a probabilistic neural network for comparison. For a signal, the

Table 1. Diagnosis results

Classifier	Feature	Misclassification		
		Outer Race	Inner Race	Ball
Fusion	Energy	0/50	0/50	0/50
Method	Kurtosis	18/50	18/50	13/50
	Energy & Kurtosis	1/50	0/50	1/50
One PNN	Energy	2/50	0/50	2/50

signal energy in each best basis is concatenated into a feature vector which then constructs the training and testing datasets. The feature vector is 6-D since there are six best bases in the case study. The probabilistic neural network uses the same smoothing parameter $\sqrt{50}$ with results shown in Table 1. Two misclassifications were recorded, i.e., for outer race and ball signal faults respectively.

The evidence produced in Table 1 clearly shows that the proposed approach is effective to conduct integrated fault diagnosis. This novel method also has superior classification capability than that using a single probabilistic neural network.

5. CONCLUSIONS

This paper has presented an approach for the implementation of an integrated machine fault diagnosis procedure based on vibration signals alone. Local decisions are made from best basis of signals' wavelet packet transform. The tasks of signal processing and feature extraction, local decision making and decision fusion are covered under one framework.

Probabilistic neural networks were used to classify features extracted from each best basis. It was shown the PNN accurately diagnosed faults in situations where relatively few training vectors were available. The weighted average and winner-take-all principles when applied in the case study were also shown to be effective for decision fusion. Signal energy as a feature extraction parameter was a good choice in bearing fault diagnosis. Poor results were obtained when kurtosis was used.

The fused decisions show that the proposed novel approach achieved higher diagnosis accuracy than a single probabilistic neural network based diagnosis.

REFERENCES

[1] I. Daubechies 1992 *CBMS-NSF Regional Conference Series in Applied Mathematics* **61** SIAM, Philadelphia, PA. Ten lectures on wavelets.

[2] S. Mallat 1989 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 674-692. A theory for multiresolution signal decomposition: The wavelet representation.

[3] B. K. Alsberg, A. M. Woodward and D. B. Kell 1997 *Chemometrics and Intelligent Laboratory Systems* **37**, 215-239. An introduction to wavelet transforms for chemometricians: a time-frequency approach.

[4] S. K. Goumas, M. E. Zervakis and G. S. Stavrakakis 2002 *IEEE Transactions on Instrumentation and Measurement* **51**(3), 497-508. Classification of washing machines vibration signals using discrete wavelet analysis for feature extraction.

[5] N. G. Nikolaou and I. A. Antoniadis 2002 *NDT&E International* **35**, 197-205. Rolling element bearing fault diagnosis using wavelet packets.

[6] D. E. Hershberger and H. Kargupta 2001 *Journal of Parallel and Distributed Computing* **61**, 372-400. Distributed multivariate regression using wavelet-based collective data mining.

[7] B. Walczak and D. L. Massart 1997 *Chemometrics and Intelligent Laboratory Systems* **38**, 39-50. Wavelet packet transform applied to a set of signals: A new approach to the best-basis selection.

[8] N. Saito, R. R. Coifman, F. B. Geshwind and F. Warner 2002 *Pattern Recognition* **35**, 2841-2852. Discriminant feature extraction using empirical probability density estimation and a local basis library.

[9] Y. Wu and R. Du 1996 *Mechanical Systems and Signal Processing* **10**(1), 29-53. Feature extraction and assessment using wavelet packets for monitoring of machining processes.

[10] S. Zhang, J. Mathew and L. Ma 2003 *Proceeding of the 10th Asia-Pacific Vibration Conference*, 835-840. Common best basis selection of wavelet packets for machine fault diagnosis.

[11] K. Mehmed 2002 *Data Mining: Concepts, Models, Methods and Algorithms*. Wiley-IEEE Press.

[12] C. M. Bishop 1995 *Neural Networks for Pattern Recognition*. New York: Oxford University Press.

[13] D.F. Specht 1990 *Neural Networks* **3**(1), 109-118. Probabilistic neural networks.

[14] S. Chen, Y. Wu and B. L. Luk 1999 *IEEE Transactions on Neural networks* **10**, 1239-1243. Combined genetic algorithm optimization and regularized orthogonal least squares learning for radial basis function networks.

[15] K. Z. Mao, K. C. Tan and W. Ser 2000 *IEEE Transaction on Neural Networks* **11**(4), 1009-1016. Probabilistic neural network structure determination for pattern classification.

[16] J. Kittler, M. Hatef, R. P. W. Duin and J. Matas 1998 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(3), 226-239. On combining classifiers.

[17] D. L. Hall and J. Llinas 2001 *Handbook of Multisensor Data Fusion*. CRC Press.

[18] http://www.eecs.cwru.edu/laboratory/bearing/download.htm.