Wavelet-Network for Classification of Induction Machine Faults Using Optimal Time-Frequency Representations

Tahar Boukra[1], Adessalam Lebaroud[1], Ammar Medoued[1]

[1] Electrical Engineering Department, University 20th august 1955, Skikda 2100, Algeria
boukratahar@hotmail.com, lebaroud@yahoo.fr, amedoud@yahoo.fr

Abstract

This paper presents a new diagnosis method for classifying current waveform events that are related to a variety of induction machine faults. The method is composed of two sequential processes: feature extraction and classification. The essence of the feature extraction is to project a faulty machine signal onto a low dimension time-frequency representation (TFR), which is deliberately designed for maximizing the separability between classes. A distinct TFR is designed for each class. The performance of fault classification is presented using two types of classifiers namely the Wavelet Neural Network (WNN) and the classical Artificial Neural Network (ANN) with Levenberg Marquardt algorithm. The flexibility of this method allows an accurate classification independently from the level of load. This method has been validated on a 5.5-kW induction motor test bench.

1. Introduction

There is a considerable demand to reduce maintenance costs and prevent unscheduled downtimes for electrical drive systems. Recencement of electrical machine faults reveals that, bearing faults, stator faults, and broken bars are the most prevalent, although almost 40%–50% of all failures are related to bearing. Bearing faults might manifest themselves as rotor asymmetry faults [1] which are usually covered under the category of eccentricity related faults. Stator faults are usually related to insulation failure; they manifest themselves through phase-to-ground connections or phase-to-phase faults. Also, stator faults manifest themselves by abnormal connection of the stator windings. Almost 30%–40% of all reported induction motor failures fall in this category [2]. The rotor fault accounts for around 5%–10% of the total induction rotor failures [2]. The principal reasons for rotor bar and ring breakage are thermal stresses due to thermal overload and unbalance. On the other side, magnetic stresses are caused by electromagnetic forces, unbalanced magnetic pull, electromagnetic noise, and vibration.

In recent years, many research works have been carried out for the study and development of fault detection and diagnosis methods of electric machines. Recent advances of signal processing techniques, such as artificial neural networks [3], wavelets [4], etc., have provided more powerful tools for fault diagnosis.

Generally, system diagnosis uses signals either in time or frequency domain. In our approach, it is potentially more informative to join both time and frequency.

Usually, the objective of time–frequency research is to create a function that will describe the energy density of a signal simultaneously in time and frequency. For explicit classification, it is not necessarily to accurately represent the energy distribution of a signal in time and frequency. In fact, such a representation may conflict with the goal of classification (i.e. generating a TFR that maximizes the separability of TFRs from different classes). It may be advantageous to design TFRs that specifically highlight differences between classes [5]-[8].

TFRs can be uniquely characterized by an underlying function called a kernel. In previous time–frequency research, kernels have been derived in order to fulfill some properties, such as minimizing quadratic interference, although some of the resulting TFRs can offer advantages for classification of certain types of signals. The goal of sensitive detection or accurate classification is rarely an explicit goal of kernel design. Those few methods that optimize the kernel for classification purposes constrain the form of the kernel to predefined parametric functions with symmetries that can not be suitable to detection or classification [8]. For classification, the optimization procedure of TFR via parameter kernel is very computationally prohibitive. It would be better to use the optimal TFR that can be classified directly in the ambiguity plane. We propose to design and use the classifier directly in the ambiguity Doppler delay plane. Since all TFRs can be derived from the ambiguity plane, no *a priori* assumption is made about the smoothing required for accurate classification. Thus, the smoothing quadratic TFRs retain only the information that is essential for classification.

In this paper, we have proposed a classification procedure based on the design of optimized TFR from a time–frequency ambiguity plane in order to extract the feature vector. We have used two classifiers, namely the wavelet neural network WNN and the classical artificial neural network ANN with Levenberg Marquardt algorithm. We gave a theoretical background for the wavelet neural network, then we presented its performance compared to the ANN. In this study, the goal is to realize an accurate diagnosis system of motor faults such as bearing faults, stator faults, and broken bars rotor faults independent from the level of load.

2. Classification Algorithm

The classification algorithm is composed of the following two parts: extraction features and decision criteria. The details of each step are described in the following paragraphs.

2.1. Feature Extraction

Here in this section, we briefly present the theory of the TFR. The class-dependent TFR [5] is defined by

$$TFR[n,k] = F_{n\to\eta}^{-1}\{F_{\tau\to k}\{\varphi[\eta,\tau]A[\eta,\tau]\}\} \quad (1)$$

where *F* represents the Fourier transform, *F-1* represents the inverse Fourier transform, η represents the discrete frequency shift, τ represents the discrete time lag, *n* represents the sample and *k* is the discrete frequency.

The characteristic function for each TFR is defined by $A(\eta,\tau)$ $\varphi(\eta,\tau)$. In other words, for a given signal, a TFR can be uniquely mapped from a kernel. The classification-optimal TFRi can be obtained by smoothing the ambiguity plane with an appropriate kernel φopt, which is a classification optimal kernel. The problem of designing the classification-optimal TFRi becomes equivalent to designing the classification-optimal kernel $\varphi opt(\eta,\tau)$.

This method, used to design kernels and thus TFRs, optimizes the discrimination between predefined sets of classes. The resulting kernels are not restricted to any predefined function but are rather arbitrary in shape. This approach ascertains the necessary smoothing to achieve the best extraction features.

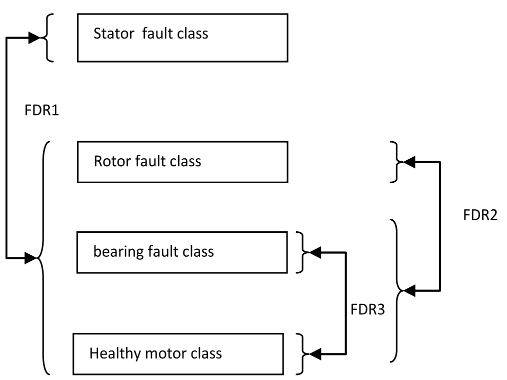

Fig. 1. Algorithm of separation between classes

Features can be extracted directly from $A(\eta,\tau).\varphi_{opt}(\eta,\tau)$ instead of the classification-optimal TFR*i*. This shortcut simplifies the computation complexity of feature extraction by reducing greatly the calculations. For further details, we recommend the reader to review our previous work [9].

Each kernel $\varphi_{opt}(\eta,\tau)$ is designed for each specific classification task. In our classification procedure, $C-1$ kernels must be designed for a *C*-class classification system. The discrimination between different classes [5] is made by separating the class i from all remaining classes $\{i+1,\ldots,N\}$ (Fig.1). In this case, the stator fault kernel is designed to discriminate the stator fault class from the other classes such as rotor fault, bearing fault, and healthy motor. The rotor fault kernel is designed to discriminate the rotor fault class from the remaining classes such as bearing fault and healthy motor. The bearing fault kernel is designed discriminate the bearing fault class from the healthy motor class. The advantage of the method lies in the optimum separation between the different classes.

For the design of any kernel, we have to determine *N* locations from the ambiguity plane, in such a way that the values in these locations are very similar for signals from the same class, while they vary significantly for signals from different classes. For that task, we have used Fisher's discriminant ratio (FDR) to get those *N* locations.

The kernels are designed by *I* training example signals from each class with the equation as follows:

$$FDR_i(\eta,\tau) = \frac{(m_i[\eta,\tau]-m_{i-remain}[\eta,\tau])^2}{V_i^2[\eta,\tau]+V_{i-remain}^2[\eta,\tau]} \quad (2)$$

where $m_i[\eta,\tau]$ and $m_{i-remain}[\eta,\tau]$ represent two means of location (η,τ)., $V_i^2[\eta,\tau]$ and $V_{i-remain}^2[\eta,\tau]$ two variances of location (η,τ).

We transform the FDR to φ_{opt} kernel in a binary matrix by replacing the maximum *N* points with one and the other points with zero. The Features can be extracted directly from $A(\eta,\tau)o\varphi(\eta,\tau)$, where *o* is an element-by-element matrix product. By multiplying the φ_{opt} kernel with a given signal ambiguity plane, we will find *N* feature points for this signal. We rank order them into a vector in order to create the training feature vector $FV_{train}^{(c)}(k)$ of class *C*

$$\varphi_{opt}^{(c)}[\eta,\tau]o\overline{A}^{(c)}[\eta,\tau] = \begin{cases} \overline{A}^{(c)}[\eta,\tau] & \text{if } \varphi_{opt}^{(c)}[\eta,\tau]=1 \\ 0 & \varphi_{opt}^{(c)}[\eta,\tau]=0 \end{cases} \quad (3)$$

where

$\varphi_{opt}^{(c)}[\eta,\tau]$, training optimal kernel;

$\overline{A}^{(c)}[\eta,\tau]$, mean class of the ambiguity plane.

Because the ambiguity plane of a signal is symmetric according to two dimensions, only points on a quarter plane are considered.

The factor of correlation is also considered. In this case, only one point is selected as a feature among a group of points that are highly correlated. These highly correlated points do not contribute to classification [10].This suboptimal approach improves the overall classification performance compared with FDR alone.

2.2. Classification Using Artificial Neural Networks

2.2.1. Artificial Neural Networks

ANNs are particularly useful for complex pattern recognition and classification tasks. The capability of learning from examples, the ability to reproduce arbitrary non-linear functions of input, and the highly parallel and regular structure of ANN make them especially suitable for pattern classification tasks [11,12].

ANNs are widely used in modeling, data analysis and diagnostic classification [13]. The most frequently used training algorithm in classification problems is the back-propagation (BP) algorithm, which is used in this work also.

There are many different types and architectures of neural networks varying fundamentally in the way they learn; the details of which are well documented in the literature [13]. In this paper, two neural networks relevant to the application being considered, i.e., classification of induction machine faults data will be employed for designing classifiers; namely the feed-forward neural network (FFNN) and the wavelet neural network WNN.

Each of the three feed-forward artificial neural network (FFNN) classifiers adopted in this algorithm for the classification of three fault types, has three layers (input-hidden-output). Extensive classification experiments were conducted to determine the optimized neural network structures. The structure of the FFNN for discriminating stator fault is 16-5-1 (input layer node number-hidden layer node number-output layer node number); the one for rotor fault is 15-6-1; the one for bearing fault is 15-6-1. The transfer and training functions adopted for the FFNN include: the hyperbolic tangent sigmoid transfer function as the transfer function for the hidden layer, the linear transfer function as the transfer function for the output layer, the Levenberg-Marquardt back-propagation as the network training function, the gradient descent learning function as the weight learning function, and the mean squared error function as the performance evaluation function. The inputs to the ANN are the normalized feature values and the output of the ANN is the binary decision made (Fig.5).

2.2.2. Wavelet Neural Networks

In this paper, a WNN was designed with one hidden layer forward neural network with its node activation function based on the so-called the Mexican hat wavelet basic function.

$$\psi(x) = \left(d - \|X\|^2\right)e^{-\frac{\|X\|^2}{2}}$$

$\|X\|$ denotes the Euclidean norm of X and $d = dim(X)$.

The applications of WNN are usually limited to problems of small input dimension. The main reason is that they are composed of regularly dilated and translated wavelets. The number of wavelets in the WNNs drastically increases with the dimension [17].

Wavelet network training consists in minimizing the usual least-squares cost function

$$J(\theta) = \frac{1}{2}\sum_{n=1}^{N}\left(y_p^n - y^n\right)^2 \tag{4}$$

where vector $\boldsymbol{\theta}$ includes all network parameters to be estimated: translations, dilations, weights of the connections between wavelets and output and weights of the direct connections; N is the number of elements of the training set, y_p^n is the output of the process for example n and y^n is the corresponding network output.

The discretized version of the wavelet is

$$\psi_{m,n}(x) = 2^{m/2}\psi(2^m x - n) \tag{5}$$

The gradient-based techniques cannot be used to adjust wavelet parameters.

Following [17], constructing a WNN involves two stages: Firstly, construct a wavelet library W of discretely dilated and translated versions of wavelet mother function ψ :

$$W = \{\psi_i : \psi_i(x) = \alpha_i\psi(\alpha_i(x - t_i))\}, \tag{6}$$

$$\alpha_i = \left(\sum_{k=1}^{N}\left[\psi(\alpha_i(x_k - t_i))\right]^2\right)^{-1/2}, \; i = 1\ldots L$$

where x_k is the sampled input, and L is the number of wavelets in W. Then select the best M wavelets based on the training data from wavelet library W, in order to build the regression

$$f_M(x) = \sum_{i\in I} u_i\psi_i(x), \tag{7}$$

where I is an M-element subset of the index set {1, 2, …, L} and $M \leq L$.

Secondly, minimize the cost function

$$J(I) = \min_{u_i, i\in I}\frac{1}{N}\sum_{k=1}^{N}\left(y_k - \sum_{i\in I}u_i\psi_i(x)\right)^2 \tag{8}$$

In principle such a selection can be performed by examining all the M-elements subsets of W. Some suboptimal and heuristic solutions have to be considered. In the following, we propose to apply two of such heuristic procedures.

The Residual based Selection

The idea of this method is to select, for the first stage, the wavelets in W that best fit the training data O_1^N , and then iteratively select the wavelet that best fit the residual of the fitting of the previous stage.

Define the initial residual $\gamma_0(k) = y_k$, k = 1,…,N, with y_k the output observations in O_1^N .

Set $f_0(x) \equiv 0$; at stage i(i=1,…,M), search among W the wavelet ψ_j that minimizes:

$$J(\psi_j) = \frac{1}{N}\sum_{k=1}^{N}\left(\gamma_{i-1}(k) - \sum_{i\in I}u_j\psi_j(x_k)\right)^2, \tag{9}$$

where

$$u_j = \left(\sum_{k=1}^{N}\left(\psi_j(x_k)\right)^2\right)^{-1}\sum_{k=1}^{N}\psi_j(x_k)\gamma_{i-1}(k), \tag{10}$$

and $\gamma_{i-1}(k)$ (k = 1,…,N) are the residual of stage i-1.

Note $l_i = \arg\min_{1\leq j\leq L} J(\psi_j)$

then ψ_{li} is the wavelet selected at stage i. Update f_i and γ_i:

$$f_i(x) = f_{i-1}(x) + u_{li}\psi(x_k)$$

$$\gamma_i(k) = \gamma_{i-1}(k) - u_{li}\psi(x_k), \quad k = 1\ldots N \tag{11}$$

Stepwise Selection by Orthogonalization

The above residual based selection procedure [16,17] does not explicitly consider the interaction or the non orthogonality of the wavelets in W. The idea of this alternative method is to select, for the first stage, the wavelet in W that best fits the training data O_1^N , and then iteratively select the wavelet that best fits O_1^N while working together with the previously selected wavelets.

The number of wavelets, M, is chosen as the minimum of the so called Akaike's final prediction error criterion (FPE) [17]:

$$J_{FPE}\left(\hat{f}\right)=\frac{1+n_{pa}/N}{1-n_{pa}/N}\frac{1}{2N}\sum_{k=1}^{N}\left(\hat{f}(x_k)-y_k\right)^2, \qquad (12)$$

where *npa* is the number of parameters in the estimator.

$n_{pa} = h(d+2)+d+1$

h is the number of wavelets in the network, d is the dimension of input vector.

After the initial WNN is constructed, it is further trained by the gradient descent algorithms like least mean squares (LMS) to minimize the mean-squared error (*mse*):

$$J(w)=\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{y}(w)\right)^2, \qquad (13)$$

Where $\hat{y}(w)$ is the real output from a trained WNN at the fixed weight vector w.

3. Experiment Results

The experimental bench consists of a three-phase asynchronous-motor squirrel cage Leroy Somer LS 132S, IP 55, Class F, T ◦C standard = 40 ◦C. The motor is loaded by a powder brake. Its maximum torque (100 Nm) is reached at rated speed. This brake is sized to dissipate a maximum power of 5 kW. Fig.2 shows the motor bench.

The wear obtained on the bearings is a real one the bearings have been provided by SECCO (Fig.3). For the rotor fault, the bar has been broken by drilling the bar of the cage squirrel (Fig.4). The 10% of power imbalance for simulating the fault of imbalance stator is obtained with a variable auto-transformer placed on a phase of the network (Fig.3). An acquisition of current signals was carried out on a test bench. The sampling rate is 20 kHz. The number of samples per signal rises to N = 100000 samples on an acquisition period of 5s. The data acquisition set consists of 15 examples of stator current recorded on different levels of load (0%, 25%, 50%, 75%, and 100%). Different operating conditions from the machine were considered, namely, healthy, bearing fault, stator fault, and rotor fault. The training set is carried out on the first ten current examples. The last five current examples are used to test the classification.

3.1. Training Set

To overcome the load problem, each class of the training set for the three faults and for the healthy machine is made of ten examples of no-load current and ten other examples for the full load. Consequently, we have 20 examples of training for each of the three faults and 20 examples of training for the healthy machine. Each signal is passed throw a low-pass filter and re-sampled with a down-sampling rate of 50. Only the range of the required frequencies is preserved. By down-sampling, the signal dimension has been reduced greatly, and using a low-pass filter is to avoid aliasing during down-sampling. The dimension of ambiguity plane is (200× 200 = 40000) points; by considering symmetry compared to the origin, we retain only the quarter of ambiguity plane, which corresponds to N = 10000.

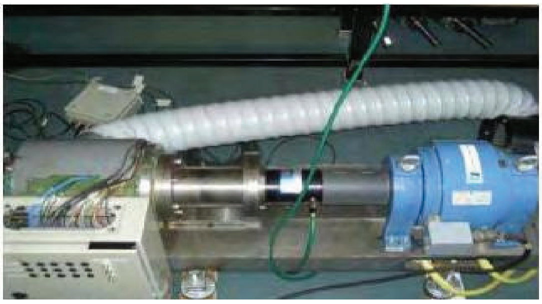

Fig. 2. Test bench of the induction motor

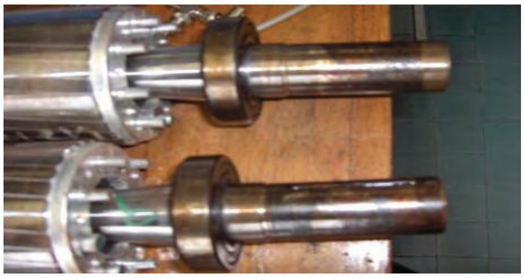

Fig. 3. Bearing faults

Fig. 4. Rotor with one broken bar.

We have tested the signals which have not been classified in the training set of the following three faults (bearing fault, stator fault, and rotor fault) with various levels of load (25%, 50%, and 75%). Five signal examples are taken for each fault and for each load level. Thus, we will have 15 signal tests for each fault.

3.2. Decision by Neural Networks

According to the theory, if the number of nodes in the hidden layer of the network is too small, the WNN may not reflect the complex function relationship between the input data and the output value. In contrast, a large number may create a complex network that might lead to a very large output error caused by over-fitting of the training sample. It was noticed that the best performance was obtained for the training set with those models whose hidden layer had 5 neurons or more Fig.6.

Against a few studies on the classification by wavelet neural network which have shown some improvement in the performance, our study demonstrates that this is not always as hollowing out. The results presented in table 1 show that the ANN with Levenberg Marquardt algorithm has a clear superiority over the latter this is due mainly to the large size of the input vector (up to 15) because WNN are usually limited to problems of small input

Table 1. Misclassification results with different classifiers:

	Kernel1	Kernel2	Kernel3
FFNN	0/15	1/15	1/15
WNN	3/15	5/15	6/15

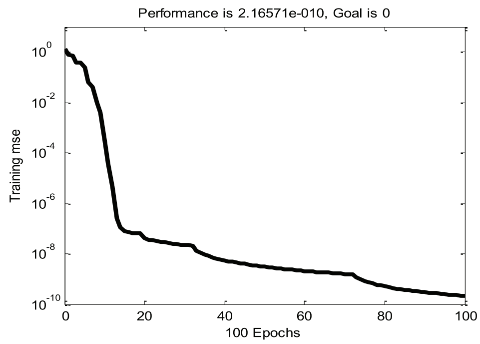

Fig. 5. Training curve for the ANN using Levenberg Marquardt algorithm

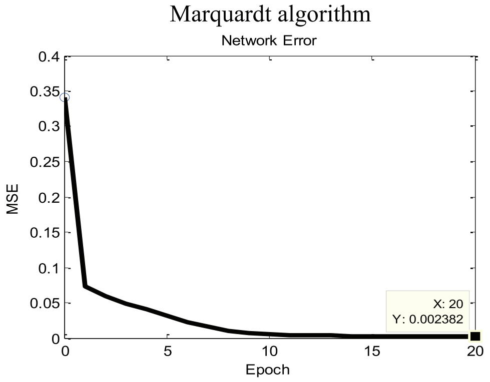

Fig. 6. Training curve for the WNN using conjugate gradient

4. Conclusion

In this paper, we have proposed a new fault classification scheme of induction machine based on TFR and criterion decision. We have based our classification on the ambiguity Doppler-delay plane where all the TFRs can be derived by a suitable choice of a kernel. Each type of fault was characterized by specific kernel. We have applied a suboptimal approach to exclude those kernel points that are strongly correlated with higher ranked kernel points. In the decision stage, the results obtained from the neural network classifiers based on wavelet or sigmoid as transfer functions, show that the ANN has a clear superiority over the WNN. These results verify that the classification scheme tested with experimental data collected from the stator current measurement at different loads, and with ANN as decision criteria is able to detect and diagnose faults with high accuracy, independently of the load condition and the type of the fault..

5. References

[1] G. B. Kliman and J. Stein, “Induction Motor Fault Detection via Passive Current Monitoring”, in *Proc. Int. Conf. Elect. Mach.*, Cambridge, MA, 1990, pp. 13–17.

[2] G. B. Kliman, W. J. Premerlani, R. A. Koegl, and D. Hoeweler, “A New Approach to On-line Turn Fault Detection in AC Motors”, in *Conf. Rec. IEEE IAS Annu. Meeting*, San Diego, CA, 1996, pp. 687–693.

[3] H. Su and K. T. Chong, “Induction Machine Condition Monitoring Using Neural Network Modeling”, *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 241–249, Feb. 2007.

[4] A. Ordaz-Moreno, R. de Jesus Romero-Troncoso, J. A. Vite-Frias, J. R. Rivera-Gillen, and A. Garcia-Perez, “Automatic Online Diagnosis Algorithm for Broken-Bar Detection on Induction Motors Based on Discrete Wavelet Transform for FPGA Implementation” , *IEEE Trans. Ind. Electron.*, vol. 55, no. 5, pp. 2193–2202, May 2008.

[5] M. Wang, G. I. Rowe, and A. V. Mamishev, “Classification of Power Quality Events Using Optimal Time–Frequency Representations” Part 2: Application,” *IEEE Trans. Power Del.*, vol. 19, no. 3, pp. 1496–1503, Jul. 2004.

[6] M. Davy and C. Doncarli, “Optimal kernels of time–frequency representations for signal classification,” in *Proc. IEEE-SP Int. Symp. Time- Freq. Time-Scale Anal.*, 1998, pp. 581–584.

[7] C. Heitz, “Optimum Time–Frequency Representations for the Classification and Detection of Signals,” *Appl. Signal Process.* vol. 2, no. 3, pp. 124–143, 1995.

[8] B. W. Gillespie and L. Atlas, “Optimizing Time–Frequency Kernels for Classification,” *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 485–496, Mar. 2001.

[9] A. Lebaroud and G. Clerc, “Classification of Induction Machine Faults by Optimal Time frequency Representations”, *IEEE Trans. on Industrial Electronics,* vol. 55, no. 12, december 2008

[10] S. Theodoridis, K. Koutroumbas, “Pattern Recognition”, Elsevier, 2006

[11] Fausett L. “Fundamentals of neural networks architectures, algorithms, and applications”, Englewood Cliffs, NJ: Prentice Hall; 1994.

[12] Haykin S. “Neural networks: a comprehensive foundation”. New York: Macmillan; 2nd ed. 1998.

[13] G. P. Zhang “Neural Networks for Classification: A Survey”, IEEE *Transactions On Systems, Man, and Cybernetics*—part c: applications and reviews, vol. 30, no. 4, november 2000

[14] Oussar Y, Dreyfus G. “Initialization by Selection for Wavelet Network Training”, *Neurocomputing* 2000;34:131–43.

[15] Subasi A, M. Yilmaz , H. R. Ozcalik “Classification of EEG Signals Using Wavelet Neural Network”, *Journal of Neuroscience Methods* 156 (2006) 360–367

[16] Q. Zhang and A. Benveniste, “Wavelet Networks”, *IEEE Trans. On Neural Networks* 3 (6) 889-898, 1992.

[17] Q. Zhang, “Using Wavelet Network in Nonparametric Estimation”, *IEEE Trans. on Neural Networks*, Vol. 8, No. 2, pp. 227-236, 1997.