A Sparse Infrastructure of Wavelet Network for Nonparametric Regression

Jun Zhang[1,2], Zhenghui Gu[1], Yuanqing Li[1], and Xieping Gao[3]

[1] Center for Brain-Computer Interface and Brain Information Processing
South China University of Technology, Guangzhou, 510640, China
[2] Information Engineering College
Guangdong University of Technology, Guangzhou, 510006, China
[3] Information Engineering College, Xiangtan University, Xiangtan, 411105, China

Abstract. In this paper, we propose a novel 4-layer infrastructure of wavelet network. It differs from the commonly used 3-layer wavelet networks in adaptive selection of wavelet neurons based on the input information. As a result, it not only alleviates widespread structural redundancy, but can also control the scale of problem solution to a certain extent. Based on this architecture, we build a new type of wavelet network for function learning. The experimental results demonstrate that our model is remarkably superior to two well-established 3-layer wavelet networks in terms of both speed and accuracy. Another comparison to Huang's real-time neural network shows that, at similar speed, our model achieves improvement in generalization performance. *abstract* environment.

Keywords: Wavelet network, neural network, sparse infrastructure, Regression.

1 Introduction

As is well known, neural network that has widely used iterative searching methodology in the learning algorithms is marked with the shortcoming of slow convergency speed, particularly in the case of large-scale observations and/or high accuracy expectation .Hence, the real-time handling capacity is a great challenge as far as neural network is concerned. Recently,wavelet was introduced into neural network to take the advantages of both the time-frequency property of wavelets and the effective learning mechanism of neural networks [2,4]. The resultant wavelet neural network has become a powerful tool for many applications. Typically, regarding function learning as a fundamental and vital problem in many fields, Jun Zhang built an orthogonal wavelet network in order to solve the redundancy and non-orthogonal problems of basis in Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) neural network. It ultimately leads the learning problem of model to solution to a set of linear equations. Therefore the iterative searching methodology can be avoided in the construction of model.

However, there still exist at least two problems as far as Jun Zhang's work is concerned. Firstly, in [3], Jun Zhang pointed out that the solution of a set

L. Zhang, J. Kwok, and B.-L. Lu (Eds.): ISNN 2010, Part I, LNCS 6063, pp. 347–354, 2010.

of linear equations may be computation intensive when the scale of problem gets larger. For real-time applications, computation load has to be controlled in solving the linear equations. Secondly, with respect to the whole function, we can build the orthogonal wavelet network for approximation that does not have any redundancy, just as [3] has done. Furthermore, aiming at a specific problem, we can employ certain algorithms for structural adaptation of the wavelet network in order to eliminate unnecessary wavelons. Due to the sparseness of large-dimensional training data, more specifically, wavelets whose supports do not contain any data point can be eliminated in the training process [5]. But concerning part of the function, the wavelet networks built by the above methods still have structural redundancy. That is to say, not all the wavelons are needed to handle part of the function.

Wavelet network model successfully maintains the favorable advantages of the RBF neural network structure. Therefore most scholars are concentrating on the algorithm improvement and pushing forward the study of the wavelet network theory, yet few pay adequate attention to the structural redundancy. For the first question mentioned above, more and more algorithms for wavelet networks have been proposed. While for the second question generally existing in all kinds of wavelet network models, restricted by the present structure of wavelet networks, such redundancy is unavoidable.

Inspired by the idea of Huang et al [6] about the neuron quantizer for general neural networks, in this paper, we propose a novel 4-layer wavelet network architecture for function learning that substantially differs from the previous wavelet network models adopting the RBF architecture. Based on the input, this novel architecture can adaptively select a subset of the wavelons by designing a group of wavelon selectors. Under this architecture, not only the structural redundancy widely existing in wavelet networks can be avoided as much as possible, but it can also control the number of wavelons in each calculation with a view of reducing the scale of the problem solutions. Furthermore, based on this infrastructure, we build a new wavelet networks and evaluates them with function learning. Performance comparison has been made on its generalization ability and learning speed with the aforementioned existing models in [2,4].

2 Orthogonal 3-Layer Wavelet Network Infrastructure

2.1 Wavelet Network for Function Learning

The wavelet decomposition of a function $f(x)$ can be expressed as

$$f(x) = \sum_{n} \langle f, \varphi_{m_0,n} \rangle \varphi_{m_0,n}(x) + \sum_{m \geq m_0} \sum_{n} \langle f, \psi_{m,n} \rangle \psi_{m,n}(x) \tag{1}$$

where $\varphi(\cdot)$ and $\psi(\cdot)$ denotes the scaling function and the wavelet function respectively; $\langle \cdot, \cdot \rangle$ represents the inner product and m_0 is an arbitrary integer representing the lowest resolution or scale in the decomposition. Given a set of training data $T_N = (x_i, f(x_i))_{i=1}^{N}$, the purpose of function learning is to find

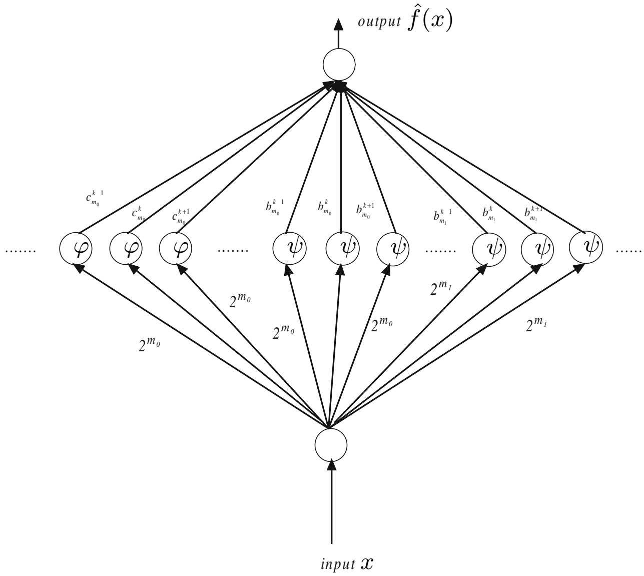

Fig. 1. The 3-layer wavelet network [5]

an estimate $\hat{f}(x)$ that closely approximates $f(x)$. For this purpose, a 3-layer wavelet network model according to the wavelet theory can be built, as shown in Fig.1. The hidden layer contains two kinds of nodes that we call φ set and ψ set both containing countable numbers of nodes. All of these nodes are named as wavelons in this paper. Lemma 2.1 gives the bias of each φ neuron and the size of φ subset. The result is the same with ψ subset.

Lemma 2.1[5]. Without loss of generality, if the support of $\varphi(x)$ is $[0,\mu]$ and that of $f(x)$ is [0,1], for the purpose of making $\{\varphi_{m_0,n} = 2^{\frac{m_0}{2}}\varphi(2^{m_0}x - n)\}$ overlay $f(x)$, the most neurons of φ set is $(2^{m_0} + \mu - 1)$ and the corresponding biases are $b_{m_0} = \{-\mu + 1, ..., 2^{m_0} - 1\}$. m_0 represents the lowest resolution or scale in the decomposition.

As a matter of fact, the problem solution can be converted to solution of a set of linear equations which has been discussed in [3]. When k is large, a direct solution to (1), which involves the inversion of a large matrix, may be computation intensive.

2.2 Shortcoming of 3-Layer Wavelet Network

As discussed above and Lemma 2.1, in the practical implementation of the wavelet network, the number of hidden layer nodes is determined by integers m_0, m_1, as well as the supports of both function f and scaling function $\varphi(\cdot)$(or wavelet function $\psi(\cdot)$). For example, suppose the interval $[0, \alpha]$ is the support of the function f and the support of scaling function is $[0,\mu]$, then the number of hidden nodes is $(2^{m_0}\alpha + \mu - 1)$ and the corresponding biases are $b_{m_0} = \{-\mu + 1, ..., 2^{m_0}\alpha - 1\}$. However, for the part of the function f whose interval is $[c, d]$,$(c > 0, d < \alpha)$, the wavelet network only need the hidden nodes whose corresponding biases are $b'_{m_0} = \{2^{m_0}c - \mu + 1, ..., 2^{m_0}d - 1\}$. So, the number of hidden nodes is $(2^{m_0}d - 1) - (2^{m_0}c - \mu + 1) + 1$. In other words, it does not entail so many hidden nodes as far as part of the function is concerned.

Without loss of generality, for d dimensional function f and distinct samples $T_N = (x_i, f(x_i))_{i=1}^{N}$, where $x_i = [x_{i1}, x_{i2}, ..., x_{id}]^T \in R^d$ and $f(x_i) \in R$, we suppose the support of f is $[0, \alpha]^d$. As far as j-th dimension is concerned, the sample set can be divided into L group as follows.

$$V_j^q = \{x_{ij} \mid \frac{(q-1)\cdot\alpha}{L} < x_{ij} < \frac{q\cdot\alpha}{L}\}, \quad 1 \le q \le L, 1 \le j \le d, q, j \in Z \tag{2}$$

and L can be set as any positive integer. Furthermore, regarding all the dimensions, we can divide the sample set into L^d groups as follows.

$$GroupV^p = \{x_i \mid \frac{(q_j-1)\cdot\alpha}{L} < x_{ij} < \frac{q_j\cdot\alpha}{L}, 1 \le j \le d, j \in Z\}, 1 \le p \le L^d, p = \sum_{j=1}^{d}(q_j - 1)\cdot L^{(j-1)} + 1 \tag{3}$$

We define a partition of function and its sub-functions as follows.

Definition 1: In view of d dimensional function f, a partition of function f is given by $f = f_1 \oplus f_2 \oplus ... \oplus f_{L^d}$, where f_i represent the i-th sub-function of f according to the supports being continuously and equally divided in all the dimensions.

Remark: According to the above discussion, we know that the data belonging to $GroupV^p$ are sampled from sub-function f_p.

Therefore, if a model can adaptively select the necessary hidden nodes to approximate the sub-function according to the input information, then the redundant wavelons for the sub-function can be eliminated accordingly. In the following section, we will introduce a novel wavelet network architecture for this purpose.

3 Proposed 4-Layer Wavelet Network Infrastructure

Assuming x as a d dimensional input, we construct the novel 4-layer wavelet network with two hidden layers and d additional wavelon selection units in the first hidden layer. As shown in Fig.3, each selection unit consists of two neurons named A-type neuron and B-type neuron. The neurons in the j-th selection unit are denoted by A_j and B_j. Regarding an arbitrary input $x \in [0, 1]$, wavelon selection unit plays the role of correctly choosing one of the L^d outputs of the second hidden layer as the final output. Such functionality is realized through choosing the weights and biases of the wavelet selection units as follows.

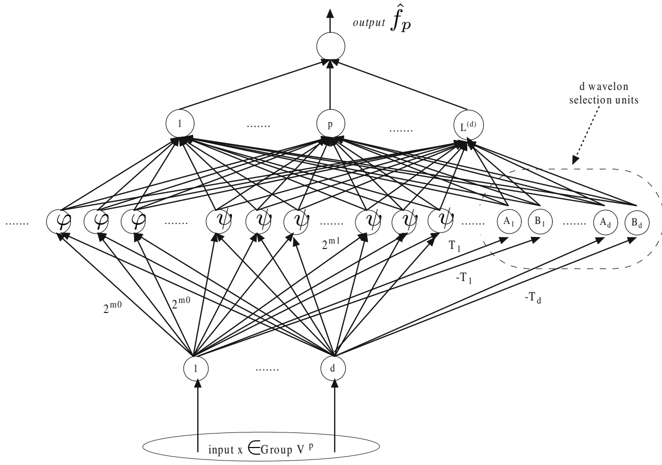

Fig. 2. Four layer architecture of wavelet network

1) The biases $b^q_{A_j}$ and $b^q_{B_j}$ of the two neurons in the j-th wavelon selection unit for the p-th neuron in the second hidden layer, where $1 \le p \le L^d, j = 1, ..., d.$, is designed as follows.

$$\begin{cases} b^q_{A_j} = -T_j(\frac{q \cdot \alpha}{L}) \\ b^q_{B_j} = T_j(\frac{(q-1) \cdot \alpha}{L}) \end{cases} \tag{4}$$

where $q = \begin{cases} \frac{p \mod L^{d-j+1} - p \mod L^{d-j}}{L^{d-j}} & j = d \\ \frac{p \mod L^{d-j+1} - p \mod L^{d-j}}{L^{d-j}} + 1 & j \neq d \end{cases}$ $1 \le j \le d$ $1 \le q \le L$.

2) Denote ω_{A_j} and ω_{B_j} as the weights that link the j-th neuron in the input layer to the j-th wavelon selection unit. Suppose $\omega_{A_j} = -\omega_{B_j} = T_j$,$1 \le j \le d$. Lemma 3.1 and Theorem 3.1 can be given as follows.

Lemma 3.1: Given arbitrarily small positive value $\eta < 0.5$, there exists a constant set

$$Tset = \{T_j^{const} | T_j^{const} = \frac{\ln(\frac{1-\eta}{\eta})}{\min_{\substack{1 \le i \le N \\ 1 \le q \le L}} (x_{ij} - \frac{q \cdot \alpha}{L})}, 1 \le j \le d\}.$$

When $T_j \ge T_j^{const}$, for $\forall x_{ij} \in V_j^k, k = 1, ..., L$, the corresponding output $Output_{A_j^q}$ of neuron A_j for the p-th neuron of the second hidden layer satisfies

$$Output_{A_j^q}(x_{ij}) = \begin{cases} \le \eta, & if\, q \le k \\ \ge 1 - \eta, & if\, q < k \end{cases}$$

Theorem 3.1[1]**:** Give arbitrarily small positive value $\eta < 0.5$, there exists a set $Tset$

$$Tset = \{T_j^{const} | T_j^{const} = \frac{\ln(\frac{1-\eta}{\eta})}{\min_{\substack{1 \le i \le n \\ 1 \le q \le L}} \left(\left| x_{i,j} - \frac{q \cdot \alpha}{L} \right| \right)}, 1 \le j \le d\}$$

when $T_j \ge T_j^{const}$,$1 \le j \le d$,for $\forall x_i \in GroupV^p, p = 1, ..., L^d$, the corresponding $Output^p_{(A_j+B_j)}(x_i)$, satisfies

$$Output^p_{(A_j+B_j)}(x_i) = \begin{cases} \le 2\eta, & x_{ij} \in V_j^q \\ \ge 1 - 2\eta, & x_{ij} \notin V_j^q \end{cases} \tag{5}$$

According to **Theorem 3.1**, $\forall x_i \in GroupV^q$,$q = 1, ..., L^d$, we can get $\eta \to 0$, thus making the inputs of the $q - th$ neuron in the second hidden layer, which come from the all of the wavelon selectors next to zero and the other neurons in the second hidden layer at least have one input that come from the wavelon selectors next to one.According to the orthogonality of wavelet, with respect to each neuron in the second hidden layer, the weights linking it to wavelons from different sub sets of φ set and ψ set can be determinated independently and successively through a transformation to the solution of a set of linear equations. The major advantage of this novel architecture is that the scale of each set of linear equations can be well controlled through function partition. On one hand, it reduces the structural redundancy as much as possibly to improve the generalization performance. On the other hand, the aim of controlling the model's computation load is achievable.

4 Experimental Results

In this section, a four-layer wavelet network (FWN) is constructed based on the novel architecture.To construct the model, we can divide the data set into two subsets, including training set $TraSet_N = (t_i, f(t_i))_{i=1}^N$ and test set $TesSet_M = (t_i, f(t_i))_{i=1}^M$. In order to evaluate the proposed architecture of wavelet networks for function learning, two examples shown in Fig.3 were given to demonstrate the validity of the presented FWN. All simulations of FWN were run in matlab 6.5 environment on an ordinary PC with Pentium 1.7 GHZ processor and 256MB memory. The measure used in [3] was taken as the performance index.

$$Error = \sqrt{\sum_{i=1}^{n} [\hat{f}(t_i) - f(t_i)]^2 / \sum_{i=1}^{n} [\bar{f} - f(t_i)]^2} \tag{6}$$

where $\bar{f} = \frac{1}{n} \sum_{i=1}^{n} f(t_i)$, and n is the number of sampled data. In Table 1, we compare the performance of our FWN model with the existing well established work in [2] [4] on the learning speed and generalization ability with the measure defined in (6). Obviously, both the two performance indexes of our model

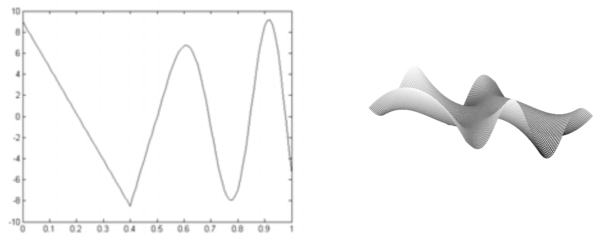

Fig. 3. Function 1 and Function 2 as examples

are remarkably superior to the previous works. Although the number of nodes has been increased, the weights and bias of most nodes can be determined at first. Because the learning algorithms of our models are not gradient-based, the approximation can be performed in a single "epoch".

Table 1. Performance comparison

Function	Model	Hid-Nodes	Epochs	Test-Err	Times(s)
Function 1	Zhang[2]	7	10000	0.0506	1100
	Pati[4]	31	800	0.0240	101.7
	BP	7	10000	0.1329	1150
	FLWN	34	1	0.0024	0.172
Function 2	Zhang[2]	49	40000	0.0339	21300
	Pati[4]	187	500	0.0230	500
	BP	225	40000	0.2938	95640
	FLWN	217	1	0.0085	3.2180

5 Conclusion

In this paper, we proposed a novel 4-layer architecture of wavelet network. Through designing a group of wavelon selection units, it adaptively selects the necessary wavelons based on the input information. The infrastructure can overcome the structural redundancy often met in conventional 3-layer architecture and bring the computation load into the effective control. Hence, it not only effectively improves the generalization performance of the model but also has faster learning capacity than the 3-layer architecture. Based on the experiment results for function learning, compared with the existing wavelet networks, the learning speed and the generalization performance of the new wavelet network models have been greatly improved.

Acknowledgments. This work was supported by the National Natural Science Funds for Distinguished Young Scholar (Grant No.60825306), Natural Science Foundation of Guangdong Province (Key Program) (Grant No.9251064101000012) and Excellent Youth Development Project of Universities in Guangdong Province (Grant No.LYM08050).

References

1. Zhang, J., Gao, X.: A new architecture of wavelet networks for function learning. IET Signal Processing (2009) (submitted to)
2. Zhang, Q., Benveniste, A.: Wavelet network. IEEE Trans. Neural Network 3, 889–898 (1992)
3. Zhang, J., et al.: Wavelet neural networks for function learning. IEEE Trans. Signal Process. 53, 1485–1497 (1995)
4. Pati, Y.C., Krishnaprasad, P.S.: Analysis and synthesis of feed-forward neural networks using discrete affine wavelet transformations. IEEE Trans. Neural Networks 4, 73–85 (1993)
5. Gao, X.P., Zhang, B.: Interval-wavelets neural networks (1)–theory and implements. Journal of software 9, 217–221 (1998)
6. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Real-Time Learning Capability of Neural Networks. IEEE Trans. Neural Networks 17, 863–878 (2006)