Wavelet Network Estimating Regression Function

Przemysław Śliwiński

Wrocław University of Technology, Institute of Engineering Cybernetics, Janiszewskiego 11/17, 50-372 Wrocław, Poland

Abstract. In the paper we propose the wavelet network estimating the regression function from the set of learning pairs. The entire framework (consisting on the network architecture, learning algorithm and pruning procedure) is derived from the theory of non-parametric estimation of the regression function.

1 Introduction

Wavelet networks are the feed-forward networks with wavelet functions as the neuron activation functions. They were primarily proposed by Zhang and Benveniste [16], however, the idea of implementing non-parametric estimation algorithms in a form of neural networks can be found in the papers of Specht [10] and [11] ; see also [1] and [7]. Wavelet networks have been already applied to recover non-linearities [8], [14], [15], estimate density functions [4], in prediction [13] and system identification [3].

In the paper we consider the properties of the wavelet network developed according to the recent results concerning the wavelet estimation of the regression function (see e.g. [2], [5], [6], [9]).

2 Theoretical background

Consider a problem of estimating a regression function $R(x)$ from the set of data $\{(x_k, y_k)\}_{k=1}^{N}$, where

$$y_k = R(x_k) + z_k,$$

under the following assumptions:

A1. Input signal x_k is stationary random white and uniformly distributed in the unit interval, $x_k \sim U[0,1]$.
A2. The function $R(x)$ is bounded in that interval.
A3. External additive random disturbance z_k is stationary white and zero-mean, independent of the input x_k.

Remark 1. The assumption **A2** is very weak and satisfied in all practical situation. Observe that $R(x)$ can be smooth or have jumps (i.e. it can be discontinuous as well).

Wavelet estimate. The wavelet estimate of $R(x)$ can have the following form

$$\hat{R}_N(x;K) = \sum_{n=0}^{2^M} \hat{\alpha}_{Mn,N} \cdot \varphi_{Mn}(x) + \sum_{m=M}^{K-1} \sum_{l=0}^{2^m} \hat{\beta}_{ml,N} \cdot \psi_{ml}(x) \qquad (1)$$

In the estimate, $\varphi_{Mn}(x)$ and $\psi_{ml}(x)$, $M, m, n, l \in \mathbf{Z}$, are scaled and translated Haar wavelet functions $\varphi(x)$ and $\psi(x)$, respectively

$$\varphi_{Mn}(x) = 2^{\frac{M}{2}} \varphi\left(2^M x - n\right) \quad \text{and} \quad \psi_{ml}(x) = 2^{\frac{m}{2}} \psi\left(2^m x - l\right)$$

and $K \in \mathbf{Z}$ is the scale parameter corresponding to the estimate accuracy; M is an arbitrary constant not greater then K. The coefficients $\hat{\alpha}_{Mn,N}$ and $\hat{\beta}_{ml,N}$ are the following sample means

$$\hat{\alpha}_{Mn,N} = \frac{1}{N} \sum_{k=1}^{N} y_k \varphi_{Mn}(x_k) \quad \text{and} \quad \hat{\beta}_{ml,N} = \frac{1}{N} \sum_{k=1}^{N} y_k \psi_{ml}(x_k). \qquad (2)$$

It was shown that for assumptions **A1-A3** (see [9]):

Lemma 1. *If the following conditions hold*

$$K = K(N) \to \infty \quad \textit{and} \quad 2^{K(N)}/N \to 0 \quad \textit{with} \quad N \to \infty \qquad (3)$$

then the estimate $\hat{R}_N(x;K(N))$ converges to $R(x)$ in the integrated mean square sense:

$$MISE\hat{R}_N(x;K(N)) = \mathrm{E} \int_0^1 \left[\hat{R}_N(x;K(N)) - R(x)\right]^2 \mathrm{d}x \to 0 \textit{ with } N \to \infty \qquad (4)$$

Moreover:

Lemma 2. *The error $MISE\hat{R}_N(x;K)$ can be disassembled into the following two parts:*

$$MISE\hat{R}_N(x;K) = O\left(2^{-K}\right) + O\left(2^K/N\right) \qquad (5)$$

The first component of (5) corresponds to the bias (systematic) error and is of deterministic nature (namely, it is the error of approximation of $R(x)$ by the wavelet expansion for the scale parameter K), the latter is the variance error and is a consequence of the randomness of the learning set.

Remark 2. The convergence (4) holds for all estimation algorithms exploiting orthogonal expansions (see [1] and [7]), however, the decomposition (5) is (under assumption **A2**) unique for wavelet-based algorithm only.

3 Wavelet network

From the Lemma 1 we conclude that the estimate (1) converges to arbitrary bounded (e.g. smooth or discontinuous) function $R(x)$. However, to assure the convergence, the complexity of the estimate (controlled by the parameter K) should grow to infinity with growing N (see (3)). Clearly, these requirements cannot be fully satisfied in practical implementations[1].

In our framework we consider wavelet network with fixed size, i.e. implementing the estimate (1) with fixed parameter K.

Network architecture. The wavelet network is the feed-forward one-and-half layer network. An example for $K = 3$ is presented on Fig. 1. For scales $m = 0, \ldots, 2$ each sum in (1) is implemented by modules of neurons with appropriate activation functions φ or ψ. The elements with labels $\hat{\alpha}_{00}$, $\hat{\beta}_{00}, \ldots, \hat{\beta}_{23}$

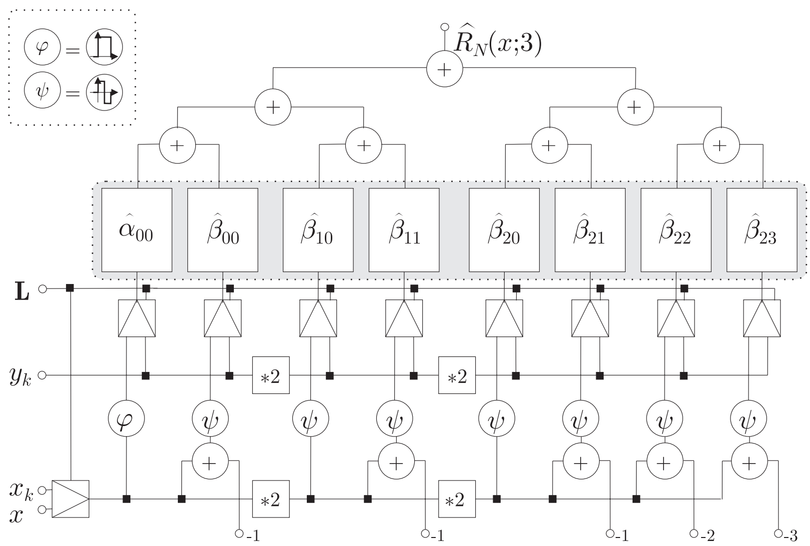

Fig. 1. Wavelet network for $K = 3$. By activating the input **L** one can learn the network *in situ*. The picture in the upper left corner illustrates that the activation functions of the wavelet neurons are extremely simple in implementation

are simple processors (arithmetic units) implementing the learning algorithm and pruning procedure described below.

[1] Hasiewicz in [5] showed that this issue can be partialy overcame and proposed *modular wavelet networks*. Like our network, they assemble modules corresponding to each value of m. However, in case of new incoming measurements, they can be enlarged by adding new modules without re-learning of the existing part.

Learning algorithm. The networks weights are equal to the coefficients $\hat{\alpha}_{Mn,N}$ and $\hat{\beta}_{ml,N}$; therefore as the learning algorithm we propose the recursive version of the equations in (2)

$$\hat{\alpha}_{Mn,N} = \frac{N-1}{N}\hat{\alpha}_{Mn,N-1} + \frac{1}{N}y_N\varphi_{Mn}(x_N)$$
$$\hat{\beta}_{ml,N} = \frac{N-1}{N}\hat{\beta}_{ml,N-1} + \frac{1}{N}y_N\psi_{ml}(x_N)$$

with the initial conditions $\hat{\alpha}_{Mn,0} = \hat{\beta}_{ml,0} = 0$.

Notice the simplicity of the algorithm. It requires only elementary arithmetic operations and has computational complexity of order $O(N)$.

Pruning procedure. The Lemma 2 says that the wavelet network of fixed size can estimate an arbitrary $R(x)$ with the finite accuracy only since for large number of learning pairs the overall error approaches the first component, independent of N

$$MISE\hat{R}_N(x;K) = O\left(2^{-K}\right) \quad \text{for large } N$$

In turn, for small sets of learning pairs, the error (5) depends only on the second component

$$MISE\hat{R}_N(x;K) = O\left(2^{K}/N\right) \quad \text{for small } N$$

Thus we can expect rather poor performance of the network for small sets of learning pairs. Exploiting once again the results of estimation theory we propose the procedure of adaptation of the network size to the number of learning pairs (i.e. a *regularization algorithm*; see e.g. [12]). Since the network is of fixed size, the procedure should prune (deactivate) these net modules for which the error introduced by the variance is greater than the reduction of the bias component (cf. (5)), i.e. for which

$$2^{-m} \leq 2^{m}/N$$

Finally, we prune (by simple zeroing of respective weights) these modules (corresponding to the scale m) for which the following inequality holds

$$N \leq 2^{2m}$$

Large size networks. For large networks (i.e. for large parameter K) one can prove the following lemma (see [9]).

Lemma 3. *If the scale parameter $K(N)$ is selected according to the rule*

$$K(N) = \left\lceil \frac{1}{2}\log_2 N \right\rceil \tag{6}$$

then the asymptotic rate of convergence (4) is of order

$$MISE\hat{R}_N(x;K(N)) = O\left(N^{-1/2}\right) \tag{7}$$

The convergence rate in (7) is guarantied for all bounded functions $R(x)$ with finite (but not known) number of discontinuities. It is worth to mention that this property of wavelet networks is unique amongst all feed-forward networks (cf. Remark 2).

4 Conclusions

The results presented in the paper draw to the conclusion that the wavelet-based estimates of the regression function can be easily implemented as a wavelet neural networks (estimates based on Haar wavelet functions are well-disposed to hardware realizations due to simplicity of activation functions). Moreover:

- The network can effectively estimate any functions met in practical applications.
- The properties of the network are independent of the distribution of the external disturbances z_k.
- The learning algorithm is simple and fast. It can be implemented within the network structure.
- The pruning procedure is automatic (controlled by the size of the learning pairs only).

References

1. K. Gałkowski and L. Rutkowski. Realizacja algorytmów rozpoznawania i nieparametrycznej estymacji za pomocą sieci neuronowych. In *Proceedings of First National Conference Neural Networks and Their Applications*, pages 237–242, Częstochowa, May 1994.
2. W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, Approximation, and Statistical Applications* . Springer-Verlag, New York, 1998.
3. Z. Hasiewicz. Non-parametric system identification by probabilistic neural network using multiresolution approach. In *Proceedings of Second National Conference Neural Networks and Their Applications*, pages 206–211, Szczyrk, 1996.
4. Z. Hasiewicz. Wavelet neural network for density estimation. In *Proceedings of Third Conference Neural Networks and Their Applications*, pages 136–141, Kule, 1997.
5. Z. Hasiewicz. Modular neural networks for non-linearity recovering by the H aar approximation. *Neural Networks* , 13, 2000.
6. Z. Hasiewicz. Non-parametric estimation of non-linearity in a cascade time series system by multiscale approximation. *Signal Processing* , 81, 2001.
7. M. Pawlak and E. Rafajłowicz. Slowly expanding orthogonal neural net for uniformly distributed input patterns — part 1 and 2. In *Proceedings of Second National Conference Neural Networks and Their Applications*, pages 365–376, Szczyrk, 1996.
8. P. Śliwiński. Self-organizing wavelet neural network – the data driven approach. In *Proceedings of Fourth Conference Neural Networks and Their Applications*, pages 153–158, Zakopane, May 1999.

9. P. Śliwiński. *Non-linear system identification by wavelets* . PhD dissertation, Wrocław University of Technology, Institute of Engineering Cybernetics, 2000.
10. D. Specht. Probabilistic neural network. *Neural Networks* , 3, 1990.
11. D. Specht. A general regression neural network. *IEEE Transactions on Neural Network* , 2, 1991.
12. V. N. Vapnik. *The Nature of Statistical Learning Theory* . Springer-Verlag, New York, 1995.
13. B. L. Zhang, R. Coggins, M. A. Jabri, D. Dersch, and B. Flower. Multiresolution forecasting for futures trading using wavelet decompositions. *IEEE Transactions on Neural Networks* , 12, 2001.
14. J. Zhang, G. Walter, Y. Miao, and W. N. Lee. Wavelet neural networks for function learning. *IEEE Transactions on Signal Processing* , 43, 1995.
15. Q. Zhang. Using wavelet network in nonparametric estimation. *IEEE Transactions on Neural Networks* , 8, 1997.
16. Q. Zhang and A. Benveniste. Wavelet networks. *IEEE Transactions on Neural Networks* , 3, 1992.