Application of Genetic Algorithms for Wavelet Networks Signal Modelling

Aleš PROCHÁZKA and Vladimír SÝS

Department of Computing and Control Engineering, Prague Institute of Chemical Technology, Technická 1905, 166 28 Prague 6, Czech Republic,
Tel/Fax: +42 [2] 332 4259 / 243 11082, E-Mail: prochaz@vscht.cz, sysv@vscht.cz

Abstract. The wavelet network has been presented recently as an efficient feedforward artificial neural network with wavelet transfer functions to solve approximation problems in nonlinear systems. The paper presents the fundamentals of wavelet networks and contributes to their analysis in the area of the initial estimate of network coefficients, taking into account the local influence of wavelets in the time and frequency domains. The main part of the paper is devoted to the study of an efficient wavelet network learning process based on genetic algorithms. The learning efficiency is assessed for different evolutions of genetic operator probabilities, and a comparison of the process is made with other approximation procedures. Results of modelling physiological signals both with wavelet and sigmoidal neural networks are presented in the final part of the paper.

1. Introduction

Nonlinear signals identification and modelling are basic tools in many engineering, biological and economical applications involving digital signal processing and time series classification and prediction. As given systems are non-stationary in most cases, various methods are studied for their analysis including the use of the wavelet transform [1, 2] and different adaptive methods.

This paper contributes to this analysis by presenting neural networks [3] with wavelet transfer functions [4, 5] instead of sigmoidal ones. The study of initialization and optimization of such wavelet networks [6, 7, 8] by genetic algorithms [9, 10, 11] is then introduced and used for electroencephalogram (EEG) signal approximation.

2. Fundamentals of Wavelet Networks

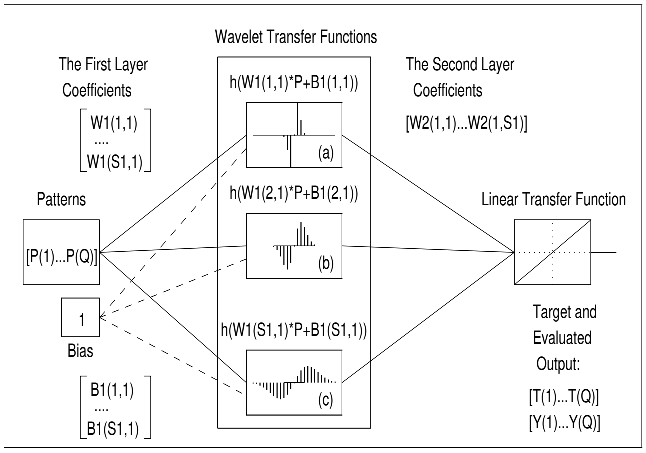
Wavelet networks are based upon the use of wavelet functions usually derived from an initial (mother) wavelet $h(x)$ dilated and translated by values forming column vectors $\mathbf{d} = [d_1, \cdots, d_{S1}]'$ and $\mathbf{t} = [t_1, \cdots, t_{S1}]'$ respectively. For a given sequence of input (pattern) values $\mathbf{P} = [P_1, \cdots, P_Q]$ it is possible to define values of individual wavelets by separate rows of the matrix

$$\mathbf{H}_{S1,Q} = \begin{bmatrix} h(\frac{1}{d_1}(\mathbf{P} - t_1)) \\ \cdots \\ h(\frac{1}{d_{S1}}(\mathbf{P} - t_{S1})) \end{bmatrix} = [h(\mathbf{W1} * \mathbf{P} + \mathbf{B1})] \quad (1)$$

for column vectors of the first layer coefficients in the form

$$\begin{aligned} \mathbf{W1}_{S1,1} &= [1/d_1, \cdots, 1/d_{S1}]' \\ \mathbf{B1}_{S1,1} &= [-t_1/d_1, \cdots, -t_{S1}/d_{S1}]' \end{aligned}$$

This set of wavelet functions can be used as a basic structure of a wavelet network as shown in Fig. 1, evaluating the



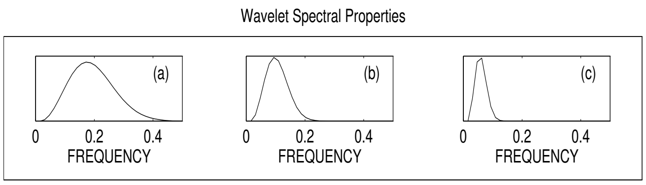


Figure 1. A two layer wavelet network of structure $1 - S1 - 1$ used for EEG signal approximation and the relation between wavelet dilation and its spectrum estimation for an initial wavelet function $h(x) = -x\ e^{-x^2/2}$

network output $\mathbf{Y} = [Y_1, \cdots, Y_Q]$ for a given pattern vector $\mathbf{P}$ and the second layer coefficients $\mathbf{W2}$ in the form

$$\mathbf{Y} = \mathbf{W2} * \mathbf{H} = \sum_{i=1}^{S1} W2(1,i)\ h(W1(i,1) * \mathbf{P} + B1(i,1)) \quad (2)$$

This structure represents a $1 - S1 - 1$ neural network using wavelet transfer functions given for a chosen initial wavelet, as shown in Fig. 1 together with their spectra esti-

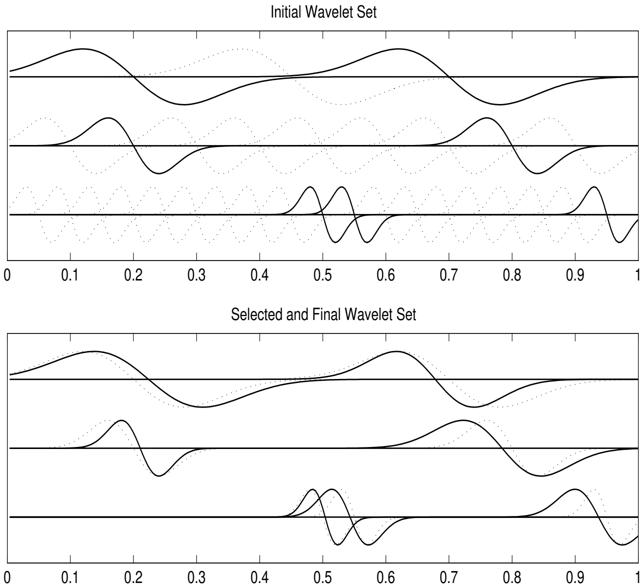


Figure 2. Initial wavelet set definition and the result of the most contributing wavelet choice for a given EEG signal approximation, shown both before and after the optimization procedure is applied

mation. In this case the optimization process results in the estimation of the dilation and translation coefficients and the evaluation of the second layer linear weights.

The theoretical background of wavelet functions has been described in various papers published recently [1, 5]. From the signal processing point of view the initial wavelet can be looked upon as a pass-band filter. Its dilation results in the corresponding pass-band compression shown in Fig. 1 for a given wavelet function.

The use of wavelet transfer functions enables their local influence to be employed both in the time and frequency domains. Even though the wavelet function theory is able to contribute to the estimation of initial coefficients [7] there are still problems with selecting a sufficiently fast and reliable optimization method since the error surface is highly nonlinear and the gradient descent method can be very inefficient.

3. Initialization and Coefficient Estimation

There are various approaches to neural network and wavelet network initialization [12]. While a random choice of coefficients usually results in a very long optimization process, the use of wavelets provides an alternative [8, 7] based on their local influence as shown in Fig. 2.

The algorithm used in this paper for wavelet network initialization for modelling a given signal consists of the following steps:

- Pyramidal wavelet structure definition using S couples of translation and dilation coefficients covering a given range of pattern values $\mathbf{P} = [P_1, \cdots, P_Q]$ and defining matrix $\mathbf{H}_{S,Q}$ by Eq.(1).
- The choice of $S1 < S$ wavelets represented by $S1$ rows of matrix $\mathbf{H}_{S,Q}$ having the largest contribution to reducing the approximation error.

To simplify the initialization process it is possible to apply the Gram-Schmidt decomposition of any matrix $\mathbf{H}_{S,Q}$ into the product of the lower triangular matrix $\mathbf{A}_{S,S}$ and the row-orthonormal matrix $\mathbf{O}_{S,Q}$ to evaluate the network error vector between the target values $\mathbf{T} = [T_1, \cdots, T_Q]$ and the network output values, using the relation

$$\mathbf{E} = \mathbf{T} - \mathbf{W2} * \mathbf{H} = \mathbf{T} - \mathbf{W2} * \mathbf{A} * \mathbf{O} = \mathbf{T} - \mathbf{V} * \mathbf{O} \quad (3)$$

for $\mathbf{V} = \mathbf{W2} * \mathbf{A}$. Defining the sum of squares in the form

$$\mathbf{E} * \mathbf{E}' = (\mathbf{T} - \mathbf{V} * \mathbf{O}) * (\mathbf{T} - \mathbf{V} * \mathbf{O})' \quad (4)$$

it is possible to evaluate subsidiary weights

$$\mathbf{V} = \mathbf{T} * \mathbf{O}' \quad (5)$$

giving the minimum of the sum of squares, with the value

$$\mathbf{E} * \mathbf{E}' = \mathbf{T} * \mathbf{T}' - \mathbf{V} * \mathbf{V}' \quad (6)$$

The choice of $S1$ fundamental wavelets can be combined with the initial matrix $\mathbf{H}$ orthonormalization. In the first step, the most contributing wavelet is found giving the maximum value of vector $\mathbf{T} * \mathbf{H}'$, with the corresponding row of matrix $\mathbf{H}$ forming the first row of matrix $\mathbf{O}$ before its normalization. In each following step that row of matrix $\mathbf{H}$ is added to matrix $\mathbf{O}$, which after its orthonormalization provides the maximum weight $\mathbf{V} = \mathbf{T} * \mathbf{O}'$. Gram-Schmidt coefficients evaluated during the orthonormalization process form the lower triangular matrix $\mathbf{A}$.

Results of the initialization process for a given EEG signal and the selection of $S1 = 7$ wavelets is shown in Fig. 2. Initial coefficients found in this way can be further optimized to minimize deviations between network output and target values in the mean square sense to achieve better results of signal approximation. Various optimization methods include genetic algorithms as well, which are especially useful for highly nonlinear systems.

4. The Use of Genetic Algorithms for Network Optimization

The principle of genetic algorithms (GA) is generally known and widely studied and used. The most famous work on the subject seems to be the Goldberg's book [10] but there are many other papers devoted to this topic published recently, providing further extensive bibliography [13, 14, 15]. We concentrate on implementation concepts here and on a brief discussion of setting parameters for the learning process.

The given time-series modelling described in the previous sections, with the one-hidden-layer perceptron-like network and wavelet activation functions, can be considered as a multi-dimensional optimization problem when one searches for both the dilations and the translations of wavelet functions which best fit the given data. These 2 $S1$ real parameters evolve through the GA and weight coefficients of the second layer are calculated by the least mean square (LMS) method. The set of dilations and translations with associated weights represents the solution of the approximation task with a certain accuracy and it corresponds to an "individual" according to the GA terminology. Particular steps of the algorithm are as follows:

1. Generation of an initial population of individuals, which means that several network sets of coefficients are chosen at random in given ranges
2. Individuals' fitness evaluation resulting from deviations between given and estimated values

Integer	*Binary code*	*Gray code*
0	000	000
1	001	001
2	010	011
3	011	010
4	100	110
5	101	111
6	110	101
7	111	100

Table 1. Comparison of binary and Gray coding

3. Parameter coding into (binary) strings of a predefined length in order to allow the use of genetic operators later
4. Random choice of individuals for the next generation according to their fitness values giving those networks with better results of data modelling a higher probability of being reproduced
5. Selection of random coefficient pairs to mate each other changing randomly their strings segments (cross-over)
6. Mutation operation turning over values of every bit in strings with relatively small probability
7. Coefficients of the current population conversion to find for each individual in the coded form its real value

In each generation the average error over the whole population can be evaluated as well as the best individual's error. The population evolves in described cycles until the solution meets a given accuracy.

The encoding is viewed as a problem which resists a more extensive application of the GA. Here two simple alternatives are employed:

- Binary coding
- Gray coding

with their codes compared in Tab. 4.

The probabilities with which the genetic algorithm operate seem to be critical for the efficiency of the learning process. Whitley [16] proposes increasing the mutation probability with time to sustain a population diversity while Etter [17] suggests reducing the mutation rate and augmenting the cross-over probability through generations. In this work three strategies are tested:

- Exponentially decreasing mutation and increasing cross-over probability
- Exponentially increasing mutation and decreasing cross-over probability
- Constant probabilities

Comparison of these choices is given in Tab. 4. The application of the first two alternatives to probabilities of mutation and cross-over provide very close results on contrary to their different strategies. The population average errors are lower in these cases with respect to that achieved by constant probabilities. The significance of binary and Gray coding seems to be not so important for the given data segment.

An example of the optimization process is shown in Fig. 3. This assumes an evolution of dilation and bias coefficients for a single element only in a given range of these coefficients, implied by properties of corresponding wavelet functions and wavelet grid density. The error surface and its contour plot illustrate the problem of its global optimum search, together with selected results of a GA application.

Method	*Binary code*	*Gray code*
Exponentially decreasing mutation and increasing cross-over	0.2233	0.1944
Exponentially increasing mutation and decreasing cross-over	0.2199	0.2149
Constant probabilities	0.3243	0.3203

Table 2. The mean values of population average errors achieved after 90 generations over 30 individuals

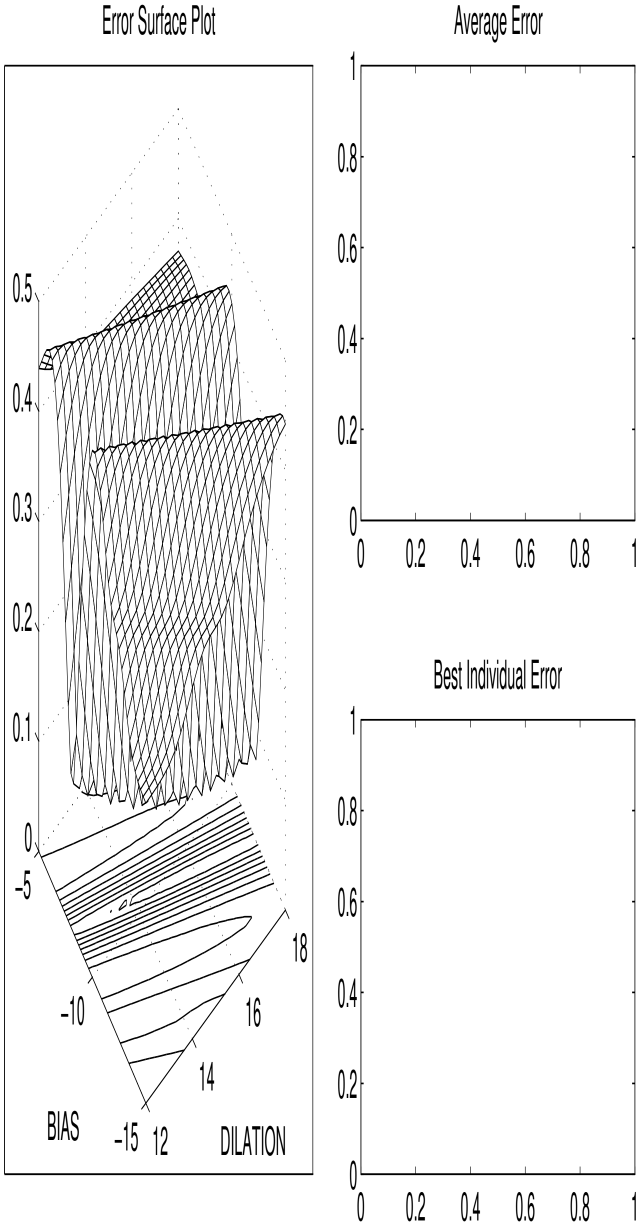


Figure 3. The genetic single element learning process for a given EEG signal approximation, together with the average error over the whole population of 16 solutions for the first 30 generations and the best individual error evolution

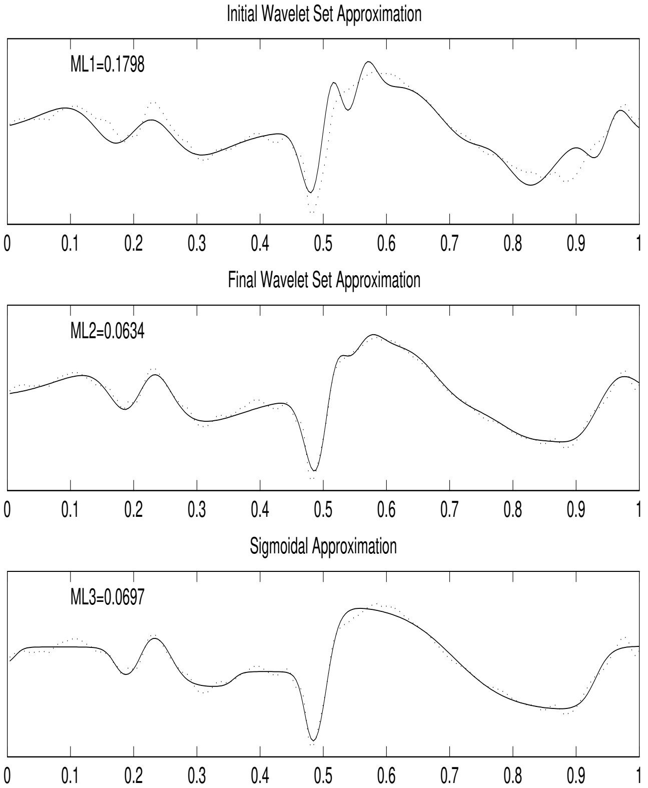


Figure 4. Result of the EEG signal approximation by $1-7-1$ wavelet network using selected wavelets before and after their coefficient optimization, compared with the sigmoidal neural network of structure $1-9-1$ approximation

5. Conclusion

This paper presents some approaches to highly nonlinear signal approximation and their verification for a given EEG segment modelling. Basic results for a given number of selected wavelets, before and after their optimization, are presented in Fig. 4. In comparison with sigmoidal neural network approximation, the number of coefficients in the wavelet case can be lower to achieve a mean square error of the same order. Using the orthonormalization procedure it is moreover possible to achieve better results of approximation by adding further wavelets, orthonormal to previous ones, with no change of their subsidiary weights.

The problem of network optimization solved in the paper is based upon genetic algorithms, showing how the development of genetic operators probabilities affect the approximation error value. Both artificial neural networks and genetic algorithms were inspired from biology, and this paper presents their joint application for signal approximation as a basic tool for signal modelling, classification and prediction.

Acknowledgements

This paper has been supported by the Department of Computing and Control Engineering at the Prague Institute of Chemical Technology. Our thanks belong also to Dr J. Jech from this Institute and to Dr J. Smith from the South Bank University in London for their valuable comments. EEG data signals were kindly provided by Dr V. Krajča from the Neurological Department of the Prague Bulovka Hospital.

References

[1] **I. Daubechies (1990)**, The Wavelet Transform, Time-Frequency Localization and Signal Analysis, *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, September 1990.

[2] **M. Unser (1993)**, On the Optimality of Ideal Filters for Pyramid and Wavelet Signal Approximation, *IEEE Transactions on Signal Processing*, vol. 41, no. 12, December 1993.

[3] **A. Lapedes and R. Farber (1987)**, Nonlinear Signal Processing Using Neural Networks: Prediction and System Modelling, LA-UR-87-2602, Tech. Rep., Los Alamos National Laboratory, USA, 1987.

[4] **O. Rioul and M. Vetterli (1991)**, Wavelets and Signal Processing, *IEEE SP Magazine*, pp. 14–38, October 1991.

[5] **M. Vetterli (1992)**, Wavelets and Filter Banks: Theory and Design, *IEEE Trans. Signal Processing*, vol. 40, no. 9, pp. 2207–2232, September 1992.

[6] **B. R. Bakshi and G. Stephanopoulis (1992)**, Wavelets as Basis Functions for Localized Learning in Multi-Resolution Hierarchy, in *Proceedings of Internationale Joint Conference on Neural Networks*, Baltimore, 1992, pp. II/140–II/145.

[7] **Q. Zhang (1993)**, Regressor Selection and Wavelet Network Construction, Tech. Rep., IRISA, France, 1993.

[8] **Q. Zhang and A. Benveniste (1992)**, Wavelet Networks, *IEEE Trans. on Neural Networks*, vol. 3, no. 7, pp. 889–899, 1992.

[9] **C. H. Chu and C. R. Chow (1993)**, A Genetic Algorithm Approach to Supervised Learning for Multilayered Networks, in *World Congress on Neural Networks*, Portland, Oregon, USA, 1993.

[10] **D. E. Goldberg (1989)**, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1989.

[11] **S. J. Marshall and R. F. Harrison (1991)**, Optimization and Training of Feedforward Neural Networks by Genetic Algorithms, in *2nd International Conference on Artificial Neural Networks*, Bournemouth, U.K., 1991.

[12] **D. Nguyen and B. Widrow (1990)**, Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights, in *Proceedings of Internationale Joint Conference on Neural Networks*, 1990, pp. III/21–III/26.

[13] **D. B. Fogel (1994)**, An Introduction to Simulated Evolutionary Optimization, *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 3–14, January 1994.

[14] **B. A. Whitehead and T. D. Choate (1994)**, Evolving Space-Filling Curves to Distribute Radial Basic Functions Over an Input Space, *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 15–23, January 1994.

[15] **J. R. McDonnell and D. Waagen (1994)**, Evolving Recurent Perceptrons for Time-Series Modeling, *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 24–38, January 1994.

[16] **D. Whitley and T. Hanson (1989)**, Optimizing Neural Networks Using Faster, More Accurate Genetic Search, in *Proc. 3rd Int. Conf. on Genetic Algorithms*, 1989.

[17] **D. M. Etter and D. C. Dayton (1983)**, Performance Characteristic of a Genetic Algorithm in Adaptive IIR Filter Design, in *Signal Processing II*, H.W.Schussler, Ed. Elsevier Science Publishers B.V. (North-Holland), 1983.