ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 345-350

Neural Network Adaptive Wavelets for Function Approximation

Fang Yong and Tommy. W. S. Chow

Department of Electronic Engineering, City University of Hong Kong,
Tat Chee Avenue, Kowloon, Hong Kong.
E-mail: eetchow@city.edu.hk

Abstract: Based on the wavelet theory, a new type of Wavelet Neural Network (WNN) is presented. For conventional neural networks (NN), the nonlinear activation function is fixed, such as the sigmodal function. In this paper, the nonlinear function is a linear combination of wavelets, that can be updated during the networks training process. This new type of WNN is applied to function approximation and it exhibits much higher learning ability compared to the conventional one. Furthermore, BP algorithm and QR decomposition based training method is derived for the proposed network.

1. Introduction

The back-propagation (BP) training algorithm is probably the most popular method used in NN, but it often suffers from getting stuck in the local minima or slow convergence. Usually, a sigmoidal function is adopted as the nonlinear activation function. Meanwhile, employing various kinds of basis functions in NN have also been investigated, such as hyper basis, splines, polynomial and radial basis function. On the other hand, wavelet theory have attracted considerable attention in many applications in signal processing and numerical analysis. Wavelet decomposition has emerged as a new powerful tool for representing nonlinearity, and a class of network combining wavelets and neural networks have recently been investigated. There are mainly two approaches to obtain the so-called wavelet neural network (WNN). One of them is to use wavelet functions as an activation functions because wavelets have many advantages over other basis functions [1,2]. The number of the hidden units was given *a priori*. The other approach is to replace each sigmoidal unit with a wavelet basis, which is determined by using the time-frequency localization properties under a given accuracy [4]. However, a relative large number of hidden units is usually required. This paper considers the combination of wavelets and neural networks for function approximation, which is similar to the first approach. The difference is that the sigmoidal activation function of NN is replaced with a linear combination of wavelet bases, and the activation function can be updated during the learning process. Basically, the activation function is adaptively determined rather than fixed. The purpose of the paper is to demonstrate the ability of function approximation of this new type of WNN. For the sake of simplicity, we consider only the one dimensional function approximation in this paper although the methods can be extended to higher dimensional case.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 345-350

2. Network architectures for approximation

Let $f(x) \in L^2(\boldsymbol{R})$ be an arbitrary function. The problem of function approximation can be stated as: given a set of training samples,

$$T_N = \{(x_i, f(x_i))\}_{i=1}^{P}, \tag{1}$$

choose the weights of a given network such that the following total squared error is minimized

$$E = \frac{1}{2}\sum_{i=1}^{P}[f(x_i) - y_i]^2, \tag{2}$$

where y_i is the actual network outputs corresponding to the ith training pattern. Here, we have assumed that the training samples were not corrupted by noise. To solve the problem, various kinds of networks have been proposed such as NN, radial basis function (RBF) network and WNN. A good network for approximation should be capable of obtaining a very small approximate error with less parameter and with a simple algorithm. Before our network is introduced, we briely review some known results.

2.1. Neural Networks

The output, $g(x)$, of a three-layer NN is represented by the following finite sums of the form:

$$g(x) = \sum_{i=1}^{N} w_i \sigma(a_i^T x + b_i), \tag{3}$$

where $w_i, b_i \in \boldsymbol{R}$, $a_i \in \boldsymbol{R}^n$, $\sigma(\cdot)$ is a given function from $\boldsymbol{R}^n$ to $\boldsymbol{R}$, $x \in \boldsymbol{R}^n$ is the input vector. It has been proved that the output, $g(x)$, is dense in the space of continuous function defined on $[0,1]^n$ if $\sigma(\cdot)$ is a continuous, discriminating function. Generally, $\sigma(\cdot)$ is adopted as a sigmoidal function that is discriminatory. However, due to the greedy nature of the BP algorithm, the training processes often settle in undesirable local minima of the error surface or converge too slowly. The purpose of this paper is to enhance the approximate capability by adaptively adopting the activation function $\sigma(\cdot)$ based on wavelet decomposition.

2.2. Wavelet neural network (WNN)

From the theory of the continuous wavelet transforms, we know that the collection of all finite sums of the form

$$g(x) = \sum_{i=1}^{N} w_i d_i^{1/2} \psi(d_i x - t_i) \tag{4}$$

is dense in $L^2(\boldsymbol{R})$. Here d_i and t_i are the dilations and translations, and $d_i > 0$. $\psi(\cdot)$ is a given function called the 'mother wavelet' whose dilation and translation

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 345-350

form a frame for the $L^2(\boldsymbol{R})$. Zhang and Benveniste ([1],1992) presented a network structure of the form

$$g(x) = \sum_{i=1}^{N} w_i \psi(d_i x - t_i) + h \tag{5}$$

where h is introduced to deal with nonzero mean functions on finite domains. Their experimental results have shown that the WNN have better approximate capability than the NN. By comparing (5) with (3), it is obvious that the sigmoidal activation function is replaced by the wavelet function.

On the other hand, the theory of the discrete wavelet transforms shows us that for some $a > 0$ and $b > 0$, the sequence $\{\psi_{mn}\}$ is a frame for $L^2(\boldsymbol{R})$ where, $\psi_{mn} = a^{n/2}\psi(a^n x - mb)$ and m and n are integers, and $\psi(\cdot)$ is a 'mother wavelet' that satisfies an admissibility condition, $\int \psi(x)dx = 0$ or $\int \left|\hat{\psi}(\omega)\right|^2 / |\omega| d\omega < \infty$. In this case, we can write an expansion of any $f(x) \in L^2(\boldsymbol{R})$ as

$$f(x) = \sum_{m,n} w_{mn} \psi_{mn} . \tag{6}$$

We call (6) the wavelet decomposition. Pati and Krishnaprasad ([3], 1993) connected the wavelet decomposition with NN by applying Daubechies' results [4]. They then proposed the following network form

$$g(x) = \sum_{(m,n)\in \mathrm{I}} w_{mn} \psi_{mn}(x) , \tag{7}$$

where the index set I is integer translation and integer dilation, which are determined by using the time-frequency localization properties under a given accuracy [3], [4], [6]. And $a = 2$, $0 < b \leq 3.5$ and $\psi(\cdot)$ is a linear combination of three sigmoids, i.e., $\psi(x) = s(x+2) - 2s(x) + s(x-2)$. In this network, only the weights w_{mn} will be identified. So it is easy to minimize E by using standard optimization algorithms and we can obtain a global minimizer w^*_{mn}. However, the number of hidden units, which is equal to the number of elements in I, is usually large.

2.3 Neural network adaptive wavelets

The choice of an appropriate activation function is crucial to the performance of the NN. In WNN, the activation function is the mother function. As wavelet decomposition has remarkable capability for representing nonlinearity, in this paper we consider that the activation function in NN is expressed by wavelet decomposition rather than fixed. This enabled the activation function to be adjusted during the learning process. The output, $g(x)$, of this new type of NN depicted in Fig. 1 is represented by the following form:

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 345-350

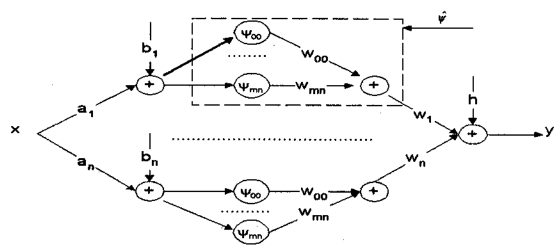

Fig. 1: The architecture of the neural network.

$$g(x) = \sum_{i=1}^{N} w_i \hat{\psi}(a_i x + b_i) + h, \tag{8}$$

where $\hat{\psi}(x) = \sum_{(m,n)\in \mathrm{I}} w_{mn}\psi(a^n x - mb)$, $\{a^{n/2}\psi(a^n x - mb) | m,n \in Z\}$ is a frame for the $L^2(\boldsymbol{R})$ and I is the index set of pairs (m,n) of integer translation and integer dilation. The parameter h is introduced so that the approximation of functions with nonzero average is possible [1]. The (8) can be rewritten as the following equivalent forms:

$$g(x) = \sum_{i=1}^{N} w_i \sum_{(m,n)\in \mathrm{I}} w_{mn}\psi[a^n(a_i x + b_i) - mb] + h \tag{9}$$

or

$$g(x) = \sum_{(m,n)\in \mathrm{I}} w_{mn} \sum_{i=1}^{N} w_i \psi[a^n(a_i x + b_i) - mb] + h \tag{10}$$

3. Simulation results

To demonstrate the approximate capability of the proposed network, it is applied to approximate two nonlinear functions. All simulations were performed under a 486DX2-66 PC by using MATLAB for Windows. In these examples, we selected the "Gaussian-derivative" $\psi(x) = -xe^{-x^2/2}$ as a mother wavelet and put $a = 2, b = 1$ [1], [4]. The error measure function E used is the mean squared error, i.e., (2). The parameter h was initialized by the mean of the function observations, and the other parameters were simply randomized between -0.5 and 0.5. It should be emphasized that the initialization of parameters is special in [3]. In this paper, we simplify the procedure of initializing the parameters in order to compare the approximation capability of the different networks.

The first example is to approximate the function $f(x) = 0.5e^{-x}\sin(6x)$ over the domain [-1,1]. The uniformly sampled test set of 100 points are available for learning. For comparison both the proposed network with 5 neurons and 9 wavelet coefficients and the wavelet network (4) with 8 wavelons are used to approximate the function with standard BP algorithm. In both cases, 25 parameters are adjusted. Fig.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 345-350

2 is the total squared errors over 2000 iterations, where the solid line shows the error of the proposed network and the dashed line represents that of the wavelet network. It is clear that the proposed network provides a better result than wavelet network. Fig. 3 shows the results with the proposed network after 5000 iterations, the total squared error is 0.0112. The solid line represents the function f while the dashed line shows the approximations.

In the above example, based on standard BP algorithm, it required 5000 iterations to converge to the total squared error of 0.0112. In order to speed up the rate of convergence and enhance the approximation capability of our network, we divide the training process into two steps. Firstly, let $\hat{\psi}(x) = \psi(x)$ and use the BP algorithm for the network training. In this case the network is a wavelet network (4) but may require less parameters. In this stage, the activation function is fixed in the training process. After some iterations, all the parameters are then fixed. The second stage is to use the QR decomposition to adjust the activation function according to the form in (10). This process is simply equivalent to solve a least squares minimisation problem. Obviously, the overall rate of convergence can be speeded up. Our results show that only 100 iterations with BP algorithm together with the QR decomposition is required to deliver a total squared error of 7.9954×10^{-5} for the same problem. Fig 4 shows the approximation result is significantly enhanced and with less computational time.

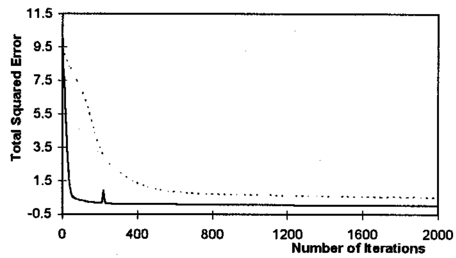

Fig. 2: The total squared errors of the WNN (4) and the proposed network for function, $f(x) = 0.5e^{-x}\sin(6x)$

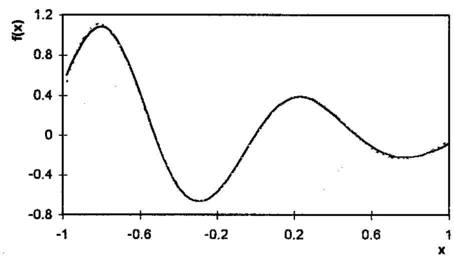

Fig. 3: The function approximation result using the BP algorithm.

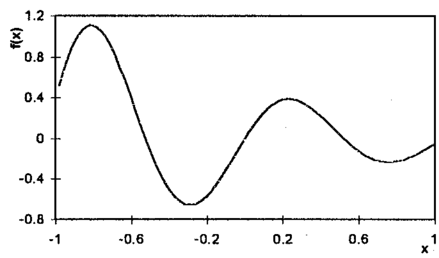

Fig. 4: The function approximation result using the proposed method.

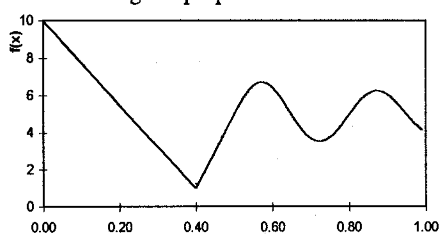

Fig. 5: The function approximation result of the second function using the proposed method.

The second example is to approximate the following function by using the proposed training method. The piecewise function is defined as

$$f(x)=\begin{cases} -22.5x+10 & 0\le x<0.4 \\ 40x-15 & 0.4\le x<0.5 \\ 5e^{-x-0.5}\sin(\frac{20\pi}{3}(x-0.5))+5 & 0.5\le x<1 \end{cases}$$

The proposed network with 5 hidden units and 25 wavelet coefficients, 41 parameters is firstly trained by the BP algorithm with 400 iterations and then QR is used. The total squared error is 0.0546. The approximate result is shown in Fig. 5.

4. Conclusion

In this paper a new type WNN is proposed for function approximation. The activation function of the WNN is a linear combination of wavelet bases, that can be updated during the training process. Simulation results are presented which indicate the network has a better approximation capability. Based on the BP algorithm and the QR decomposition, a training method for the proposed network was also derived and has been validated by the given results. The further theoretical work in exploring the selection of different wavelet bases will be reported in our later work.

5. References

[1] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. on Neural Networks*, vol. 3, no. 6, pp. 889-898, 1992.

[2] T. Yamakawa, E. Uchino and T. Samatsu, "Wavelet neural networks employing over-complete number of compactly supported non-orthogonal wavelete and their applications," *IEEE Int. Conf. on Neural Network*, vol. 1, pp. 1391-1396, 1994.

[3] Y. C. Pati and P. S. Krishnaprasad, "Analysis and synthesis of feedforward neural networks using discrete affine wavelet transformations," *IEEE Trans. on Neural Networks*, vol. 4, no. 1, pp. 73-85, 1993.

[4] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. on Information Theory*, vol. 36, no. 5, pp. 961-1005, 1990.

[5] J. Zhang, G. G. Walter, Y. Miao and W. N. Wayne, "Wavelet neural networks for function learning," *IEEE Trans. on Signal Processing*, vol. 43, no. 6, pp. 1485-1496, 1995.

[6] I. Daubechies, *Ten lectures on wavelets*, Philadelphia, PA: SIAM Press, 1992.