Adaptive Neural Activation Functions in Multiresolution Learning
Yao Liang
Alcatel USA
Raleigh, NC 27609, USA

Abstract
In this paper, we extend our original work on multiresolution learning, and present a new concept and method of adaptive neural activation functions in multiresolution learning, to maximize the learning efficacy of multiresolution learning paradigm for neural networks. Real-world sunspot series (yearly sunspot data from 1700 to 1999) prediction has been used to evaluate our method. We demonstrate that multiresolution learning with adaptive activation can further significantly improve the constructed neural network's generalization ability and robustness. Therefore, our work demonstrates the synergy effect on network learning efficacy through multiresolution learning with neural adaptive activation functions.

1 Introduction
Generalization is a key requirement and yet very difficult problem in neural network research and applications. Aimed at improving the generalization ability of neural networks for signal prediction, a new and effective learning paradigm, called multiresolution learning, has been presented [1, 2]. Unlike traditional learning paradigm, the multiresolution learning paradigm exploits the correlation structures in the training data at multiple resolutions, which otherwise could be obscured at the original resolution of the training data. It has been demonstrated that multiresolution learning can significantly improve neural network's generalization performance [1, 2] and neural network robustness [3] on difficult signal prediction tasks.

In this paper, we further explore relationship between the resolution of training data and the steepness of activation function of neurons. We investigate how to maximize the efficacy of multiresolution learning by adapting neuron's sigmoid activation functions during the multiresolution learning process. We introduce a new concept and method to adapt the slope of neural sigmoid activation functions to the training data at different resolution during multiresolution learning process. We will show how this novel and simple approach can further improve the generalization ability and robustness of the constructed neural networks, and therefore maximize the efficacy of multiresolution learning paradigm.

The rest of the paper is organized as follows. Section 2 provides a review of multiresolution learning in order to put our this work into the context. The method is presented in Section 3. In section 4, simulation results are provided to demonstrate our method. Finally, conclusion remark is given in Section 5.

2 Multiresolution Learning Review

Multiresolution learning is based on multiresolution analysis [4, 5] in wavelet theory. The multiresolution analysis framework is employed for decomposing the original signal and approximating it at different levels of detail. Unlike traditional neural network learning which employs a single signal representation for the entire training process, multiresolution learning exploits the approximation sequence representation-by-representation, from the coarsest version to finest version during the neural network training process. In this way, the original signal, the finest resolution version in the approximation sequence, will finally be used in the learning process.

Assume that a given sampled signal $s^n$ is to be learned. Let $m, M \in \mathbb{Z}$ and $0 < M < m$. Let a learning activity $A_{i}(r_j)$ denote a specific training phase conducted on the representation $r_j$ of training data (which may include some form of preprocessing) with any given learning algorithm. Let $\rightarrow$ be the learning dependency operator by which $A_{i} \rightarrow A_{j}$ means that the learning activity $A_{i}$ should be conducted before the learning activity $A_{j}$. Multiresolution learning then can be defined as a sequence of learning activities $\{A_{i}(r_j)\}_{m \geq j > M}$ associated with the sequence of approximation subspaces $\{V_j\}$ in multiresolution analysis such that the following requirements are satisfied.

1. The representation $r_j$ is associated with the approximation $s^j$ of the original signal $s^n$ in the approximation subspace $V_j$.
In this paper, we investigate how to maximize the efficacy of multiresolution learning by adapting neuron's sigmoid activation functions during the multiresolution learning process. Since training data at different resolution level can potentially expose different inherent characteristics and correlation structure, it would be desirable to employ different neural processing capability of neurons which will be more appropriate and effective for each different resolution training data. Our method is to adapt the sigmoid activation functions of neurons (in feed-forward neural networks) to the training data at different resolutions by means of adjusting the steepness of the activation functions in the region. We will show how this novel approach can improve the generalization ability and robustness of the constructed neural networks, and therefore maximize the efficacy of multiresolution learning paradigm.

A typical sigmoid activation function of neurons is the logistic function

$$f(\text{net}) = \frac{1}{1 + \exp(-\beta \text{net})},$$

where \(\text{net} = \sum_{i} w_i x_i\), the summing product of the connection weights and the corresponding neuronal outputs from units on the previous layer, and \(\beta\) is the slope parameter of the activation function \(f()\) in the region.

The sigmoid activation function (Equation 2) is shown in Figure 1 with several different slopes. In the plot, slope \(\beta\) ranges from 1.0 to 2.5, in increments of 0.5. The larger value of the parameter \(\beta\), the steeper curve of the activation function.

Our insight is that, in multiresolution learning paradigm, any incremental change of neurons' activation function input \(\Delta x_i\) would be more significant in the training with coarser resolution data than the same amount of that in the training with finer resolution data. This would suggest that network neurons in learning process should be more perceptible to any input change \(\Delta x_i\) of coarser resolution data than that of finer resolution data, in order to achieve the potential maxim learning efficacy. Based on this insight, we present a novel concept and method to adapt the neural activation functions according to the resolution level of training data in the way that steeper activation functions are used for coarser resolution training data in multiresolution learning process. A simple adapting scheme is given as follows. Assume that there are total \(K\) learning activities \(A_j (j = 1, \ldots, K)\) (i.e., total \(K\) different resolution levels of training data) in a multiresolution learning process. Let \(a, b \in \mathbb{R}\) and \(0 < a < b\), where \(a\) and \(b\) are the values of activation slope parameter \(\beta\) used for the finest and coarsest resolution versions of training data respectively. Then, the neural activation slope \(\beta\) for any \(A_j (j = 2, \ldots, K-1)\) would be adapted to a value with \((b-a)/(K-1)\) in difference of that adopted for its two adjacent learning activities. All neurons of network with same resolution level training data will have the same slope \(\beta\) for their sigmoid activation functions.

The concept and method of adaptive activation functions developed here should not be confused with the adaptive
neuron model which was studied previously at different concept and context [6, 7, 8]. In our method, the slope of sigmoid activation functions is adjusted in the context of multiresolution learning when the resolution level of the training data is changed, and is completely independent with backpropagation or any other training algorithm used, whereas in the adaptive neuron model, the slope adaptation was a part of backpropagation weight update procedure, in which the slope of each neuron is updated individually after every weight update step, to reduce the computation time of backpropagation algorithm.

4 Simulation Results

To demonstrate our method of adjusting slope of activation functions in multiresolution learning, real-world nonlinear time series prediction is investigated. Sunspot series, a challenging problem posed by nature, is a benchmark time-series forecasting problem frequently studied by time-series analysts [9] and investigated by neural network researchers [10, 11, 12, 13]. Sunspots are dark blotches on the sun, and often larger in diameter than the earth. First observed around 1610, yearly averages of sunspots have been recorded since 1700. Weigend et al. [10, 11], investigated sunspot series prediction using feedforward neural networks with weight-elimination and the resulting 12-3-1 network structure. (The commonly used notation 12-3-1 denotes a three-layered network having 12 input nodes, 3 neurons in the hidden layer, and a single output neuron.) It has been demonstrated in [1, 2] that multiresolution learning can achieve significantly better prediction performance than that attained by Weigend et al. neural network, which is considered as one of the best benchmarks for sunspot series prediction.

Previously published results have employed sunspot data between the years 1700 and 1920 to train the neural network and data from 1921 to 1979 for evaluating forecasting performance. In our following experiments, the yearly sunspot data from 1700 to 1999, the very beginning recorded data through the most recent data observed to date [14], as shown in Figure 2, is employed. The comparison simulations are conducted using multiresolution learning with and without adjusting neural activation functions. The sunspot data between years 1700 and 1939 are used as training set to construct neural networks, and the sunspot data from 1940 to 1999 are used for evaluating iterated multi-step prediction performance.

4.1 Experiments

In our experiments, the sunspot training data $s^m$ is decomposed as

$$s^m = s^{m-2} + d^{m-2} + d^{m-1}. \quad (3)$$

The Haar wavelet basis is used for the decomposition. From this decomposition, two approximation versions at coarser resolutions of training data $s^{m-2}$ and $s^{m-1} = s^{m-2} + d^{m-2}$ are obtained. The corresponding multiresolution learning process for the sunspot series then will contain $A_m(x(s)), A_m(s(m)), \text{and } A_d(s)$, where

$$r_j = \begin{cases} s^j & j = m \\ s^j + \sum_{k=j}^{m-1} d^k : d^k = 0 & j = m-2, m-1 \end{cases}. \quad (4)$$
The 12-3-1 neural network structure is used, in which hidden neurons employ sigmoid activation functions, while the output neuron employs a linear activation function. The multiresolution learning process is

\[ A_{m-2}(r_{m-2}) \rightarrow A_{m-1}(r_{m-1}) \rightarrow A_m(r_m). \]  

(5)

When using multiresolution learning without adjusting neural activation functions, each of the three learning activities \( A_{m-2}(r_{m-2}), A_{m-1}(r_{m-1}), \) and \( A_m(r_m) \) in Equation 5 was conducted with the slope parameter \( \beta \) set to 1.0. When using multiresolution learning with adjusting neural activation functions, \( \beta \) was set to 2.0 for the first learning activity \( A_{m-2}(r_{m-2}) \), and then adjusted to 1.5 for the second learning activity \( A_{m-1}(r_{m-1}) \), and then adjusted to 1.0 for the final learning activity \( A_m(r_m) \). In both cases, the training process was started with identical initial random connection weights in the backpropagation procedure, and the same learning rate = 0.01. No momentum term was used.
4.2 Results

To compare the results, the normalized mean squared error (NMSE) [15] is used to assess forecasting performance. The NMSE is computed as

\[
\text{NMSE} = \frac{1}{\sigma^2} \frac{1}{N} \sum \{x(t) - \hat{x}(t)\}^2
\]

where

\(x(t)\) is the observed value of the time series at time \(t\);
\(\hat{x}(t)\) is the predicted value of \(x(t)\); and
\(\sigma^2\) is the variance of the time series over the prediction duration.

The results of NMSE in forecasting (iterated multi-step prediction) using neural networks trained from two different sets of initial random connection weights are shown in Figures 3 and 4. In order to aid in analyzing the results, the NMSE for the traditional learning approach (with identical network structure, initial weights and learning parameters) is also given accordingly as a 'bottom line' of comparison, in Figure 3 and Figure 4.

As seen in Figures 3 and 4, the forecasting (generalization) performance of the network employing multiresolution learning with adaptive activation slope is superior to that obtained from the network employing plain multiresolution learning, which, as demonstrated before, is significantly better than that achieved from the network using traditional learning. The comparison clearly illustrates the synergy effect on improving generalization performance in multiresolution learning paradigm with adaptive neural activation slope. Furthermore, while the network performance is sensitive and moderately sensitive to the random initial connection weights employing traditional learning and plain multiresolution learning respectively, multiresolution learning with adaptive activation is able to significantly reduce the weight initialization sensitivity and achieve more robust network performance. In addition, due to the steeper neural activation adapted for coarser resolution training data, 10% speedup in learning is gained.

One of the constructed neural networks employing multiresolution learning with adaptive activation slope in our sunspot series prediction experiments (evaluated in Figure 4) is given in Appendix.

5 Conclusion Remark

In this paper, the relationship between the resolution level of training data and the steepness of sigmoid activation function of neurons is studied. A novel concept and method of adaptive neural activation slope in multiresolution learning paradigm is presented to maximize the learning efficacy. The synergy effect on improving network generalization and robustness has been demonstrated through our simulation results on real-world sunspot series predictions. We believe that this work, as an important extension to our original work on multiresolution learning [1, 2], not only provides with helpful insight on improving our understanding the nature of neural network learning, but also offers a useful and simple method in achieving better generalization performance and robustness of neural networks for difficult applications in real-world situations.

Appendix

The parameters of a constructed neural network employing multiresolution learning with adaptive activation functions for sunspot series prediction reported in this paper are listed in Table1 and Table 2 (see the next page). The original sunspot values from NOAA National Geophysical Data Center [14] were divided by 50 for scaling, and then shifted (a bias) to obtain the zero mean for the training set.

References


2605
Table 1. Connection Weights of the Constructed Neural Network

<table>
<thead>
<tr>
<th>Neuron</th>
<th>hl</th>
<th>h2</th>
<th>h3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurons to Output Neuron</td>
<td>-1.790601</td>
<td>1.090282</td>
<td>2.842443</td>
</tr>
</tbody>
</table>

Input to Neurons

<table>
<thead>
<tr>
<th>t-11</th>
<th>t-10</th>
<th>t-9</th>
<th>t-8</th>
<th>t-7</th>
<th>t-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>To hl</td>
<td>0.687223</td>
<td>-1.024738</td>
<td>-0.339208</td>
<td>0.494194</td>
<td>-0.814053</td>
</tr>
<tr>
<td>To h2</td>
<td>-0.071999</td>
<td>0.273022</td>
<td>-0.533368</td>
<td>-0.673927</td>
<td>0.894344</td>
</tr>
<tr>
<td>To h3</td>
<td>0.416260</td>
<td>-0.624725</td>
<td>-0.044752</td>
<td>0.434757</td>
<td>-0.590601</td>
</tr>
</tbody>
</table>

Input to Neurons

<table>
<thead>
<tr>
<th>t-5</th>
<th>t-4</th>
<th>t-3</th>
<th>t-2</th>
<th>t-1</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>To hl</td>
<td>-0.824582</td>
<td>-0.302282</td>
<td>-0.825485</td>
<td>-1.009792</td>
<td>1.523415</td>
</tr>
<tr>
<td>To h2</td>
<td>-1.503858</td>
<td>1.276529</td>
<td>-1.946989</td>
<td>-1.012489</td>
<td>0.587903</td>
</tr>
<tr>
<td>To h3</td>
<td>-0.130716</td>
<td>-0.455797</td>
<td>-0.369611</td>
<td>-0.387429</td>
<td>-0.044598</td>
</tr>
</tbody>
</table>

Table 2. Neuron Biases of the Constructed Neural Network

<table>
<thead>
<tr>
<th>Neuron</th>
<th>hl</th>
<th>h2</th>
<th>h3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>-1.063669</td>
<td>-0.012807</td>
<td>-1.546708</td>
</tr>
</tbody>
</table>

[14] Sunspot series can be obtained from NOAA National Geophysical Data Center website: 
http://www.ngdc.noaa.gov/

2606