International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.8 (2014), pp.149-156
http://dx.doi.org/10.14257/ijmue.2014.9.8.13

An Adaptive Stopping Criterion for Backpropagation Learning in Feedforward Neural Network

J. T. Lalis[1], B. D. Gerardo[2] and Y. Byun[3*]

[1]*College of Computer Studies, La Salle University, Ozamiz City, Philippines*
[2]*Institute of ICT, West Visayas State University, Luna St., Lapaz Iloilo City, Philippines*
[3]*Dept. of Computer Engineering, Jeju National University, Jeju, Korea*
j_lalis@yahoo.com[1], bgerardo@wvsu.edu.ph[2], and ycb@jejunu.ac.kr[3]

Abstract

In training artificial neural networks, Backpropagation has been frequently used and known to provide powerful tools for classification. Due to its capability to model linear and non-linear systems, it is widely applied to various areas, offering solutions and help to human experts. However, BP still has shortcomings and a lot of studies had already been done to overcome it. But one of the important elements of BP, the stopping criterion, was given a little attention. The Fisher's Iris data set was used to this study as input for standard BP. Three experiments, using the different training set sizes, were conducted to measure the effectiveness of the proposed stopping criterion. The accuracy of the networks, trained in different data set sizes were also tested by using the corresponding testing sets. The experiments have shown that the proposed stopping criterion enabled the network to recognize its minimum acceptable error rate allowing it to learn to its maximum potential based on the presented patterns. The ubiquitous stopping criterion presented in this paper proved that the number of iterations to train the network should not be dictated by human since the accuracy of the network depends heavily on the number and quality of the training data.

Keywords: *Artificial neural networks, Backpropagation, adaptive stopping criterion*

1. Introduction

The artificial neural networks structure mimics the brain's biological neurons in order to learn and recognize patterns. Various fields [7] have already been using the ANN to solve variety of tasks that are difficult to solve using rule-based programming. Furthermore, it has the capability to build models for both non-linear and non-linear systems even if the relationship between the input and output data is not known [1, 2].

The structure of ANN depends on a particular application. Once its structure has been defined, it is ready to be trained. This will allow the network to learn appropriate behavior based on the presented patterns. There are different paradigms to train ANN and supervised learning is one of it. In supervised learning method, each training example consists of input object and a supervisory signal. Backpropagation is usually considered as a supervised learning method. BP was first described by Arthur E. Bryson and Yu-Chi as a multi-stage dynamic system optimization method in 1969 but it did not gained any recognition. Not until 1986, the work of Geoffrey E. Hinton, David E. Rumelhart, and Ronald J. Williams led the revival of BP in the field of ANN research. The use of BP learning method to train feedforward neural networks has been proven to provide powerful tools for

* Corresponding Author

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2014 SERSC

analyzing real world and complex problems such as financial diagnosis [3], security [4], agriculture [5], handwriting recognition [8], and even diagnosis [9].

However, the BP learning method has major shortcomings such as slow learning speed associated with computational complexity and converging into local minima [6]. To address these problems, numerous studies had already been conducted to modify and improve the standard BP algorithm. Unfortunately, the stopping criterion of BP, which is considered as one of the key elements of BP, has been scarcely studied. In the reviewed literature, the stopping of training of either standard or modified BP depends on human decision. In some studies [7], the training is limited on the epochs or the number of rounds with the given patterns. But this approach may prevent the potential of neural networks to learn more and further minimize the error rate. Moreover, the risk of converging into local minima is also high in this approach. Other studies [8, 2] also used predefined acceptable error rate to halt the training. However, there is a tendency that training process will never end if the given acceptable error rate is unreachable based on the patterns presented to the ANN.

The major objective of this study is to further improve the standard backpropagation learning algorithm used in training the multilayer feedforward neural network by adding a ubiquitous stopping criterion on it. This method allows the network to learn with lesser intervention to human, wherein, acceptable error rate is being recognized by the network itself.

2. Literature Review

2.1. The Feedforward Artificial Neural Network

The feedforward artificial neural network, or also known as multilayer perceptron, is a fully connected network model that maps the input data sets into the corresponding output sets. It was the first and simplest type of neural network composed by multiple layers of nodes known as the input layer, hidden layer, and the output layer. As shown in Figure 1, each layer is connected to the next layer through its processing elements known as neurons in weighted links.

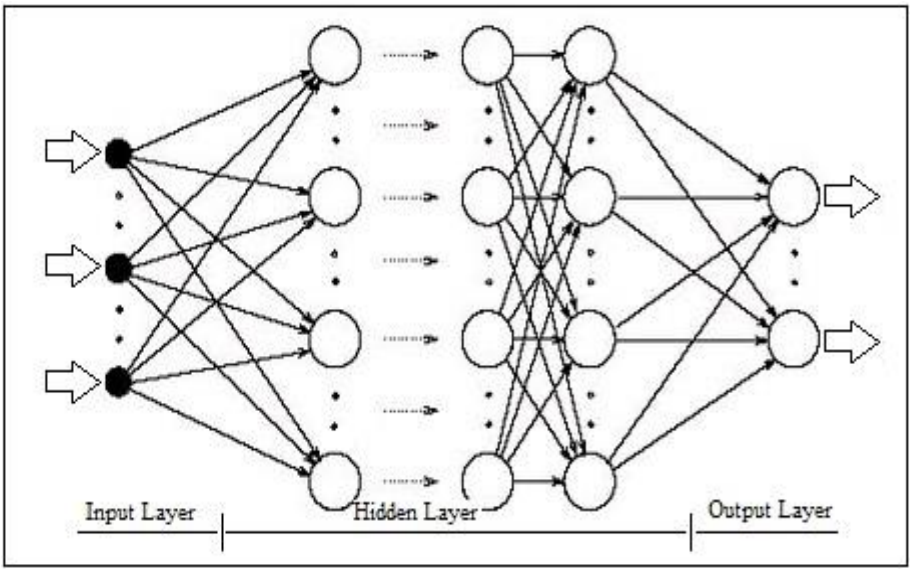

Figure 1. Feedforward Artificial Neural Network Architecture

 Copyright ⓒ 2014 SERSC

Each neuron has nonlinear activation function that scales its net input into a specific range. In the case of multiple hidden layers, the output from one hidden layer is forwarded into the following hidden layer and separate weights are provided to the sum going into each layer.

2.2. Backpropagation Learning Algorithm

Multi-layer networks use a variety of learning techniques and backpropagation is the most common learning methodology used on it. BP is also one of the simplest and most general methods for the supervised training and data mining [5]. Learning process in BP allows the neural network to produce a desired response by adapting itself to stimulus. The network adjusts its synaptic weights as it receives the input stimulus. The learning process is done continuously until the actual response converges to the expected response.

During the training session, the weights of the nodes in the hidden and output layers are initialized with small pseudo-random numbers. Then the input pattern is propagated forward and the activation level of the hidden and output nodes is calculated using an activation function. The weights are then updated starting from the output layer nodes backward to the hidden layer(s) nodes. The error gradient between the actual and expected responses is then calculated using the delta rule. This causes the neuron to produce an error signal. This error signal changes its value depending on the weights of the neurons in each layer. The error triggers the network to adjust its weights and minimize the error rate until it converges to the desired outcome.

The training process can be stopped using an error criterion. It could be either when the error is below or equal to predefined acceptable error rate; or by limiting the number of epochs. The second criterion ensures that the learning process will stop, however, there is no assurance that the network will be able to learn into its maximum potential. Furthermore, the first criterion doesn't ensure that the learning process will stop if in case that the given acceptable error rate is unreachable based on the patterns being presented on the network.

3. The Adaptive Stopping Criterion Algorithm

This study focuses on the development of an adaptive stopping criterion that will enable the standard BP algorithm to converge to its global minima. The algorithm works as follow,

1. Set the value of threshold to 0.00, counter to 0, the previous mean squared error (MSE) and least MSE to 1.
2. Initialize the predefined feedforward neural network.
3. Calculate the MSE of the network based on the presented patterns (epoch) using the standard BP learning method.
4. Round off the MSE to the nearest five decimal places.
5. If the previous MSE is equal to the MSE decrement counter by 1, otherwise, increment counter by 1.
6. Override the value of the previous MSE with the calculated MSE.
7. Assign the value of MSE to the least MSE if it is less than the least MSE.
8. Check if the value of counter is equal to 0. If it is equal to 0:

   8.1 calculate the value of the threshold with the following formula:

Copyright ⓒ 2014 SERSC

Threshold = (threshold * 3 + least MSE) / 2

8.2 Assign the calculated threshold to least MSE if it is greater than the least MSE.

8.3 If the MSE is less than threshold, stop the training, otherwise, start again from step two (2) again until the network stops learning.

9. If counter is not equal to zero (0), start again from step three (3) until the network stops learning.

A minimal change in the movement of the patterns based on the calculated delta rule is signified by the mean squared error rounded off into five decimal places.

4. Experiments and Results

To test the effectiveness of the proposed stopping criterion, the Fisher's Iris data set, has been used as input patterns to train the feedforward neural networks using the standard BP algorithm. This data set is a multivariate data set that contains three different classes (setosa, versicolor and virginica). Each class contains 50 instances or patterns that describe a certain type of Iris plant. In this data set, each class is linearly separable with each other. To avoid over-fitting, the data set was divided into two parts, namely training and testing sets. Three experiments were conducted and with different ratios of training and testing data set sizes. For the first experiment, 20% of the entire data set was used for the training and the rest was for the testing. The 40% of the data set was used to train the neural networks and the remaining 60% was used to test it in the second experiment. And lastly, 60% for training and 40% for training. The prediction accuracy of the trained ANN using the BP with adaptive stopping criterion in each experiment was then recorded and analyzed.

4.1. Experiment 1

The y axis in Figure 3 shows the mean squared error while the number of epochs appears on the x axis. It may be seen clearly that the threshold remains stable at zero. The sudden movements of MSE trigger the threshold to increase gradually and reach its peak causing the network to stop the learning process at 4,192nd epoch.

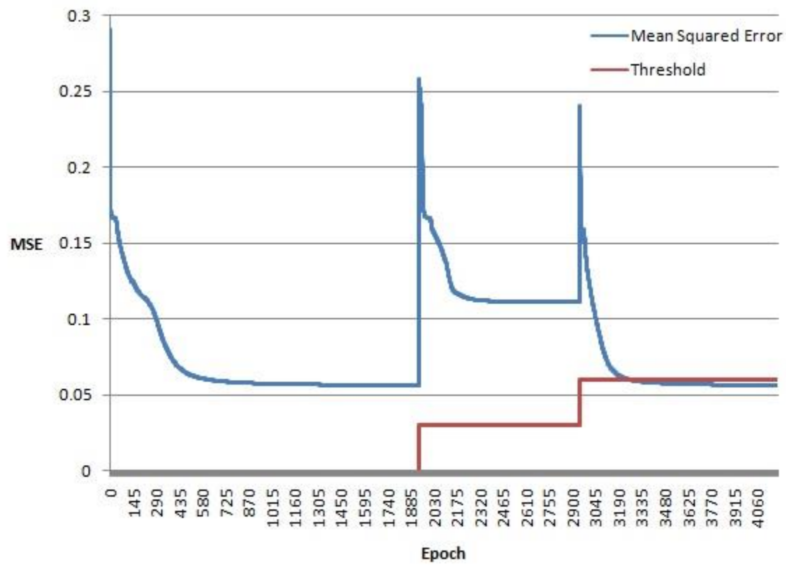

Figure 2. Convergence at the Global with 20% of Iris Data Set

Copyright ⓒ 2014 SERSC

4.2. Experiment 2

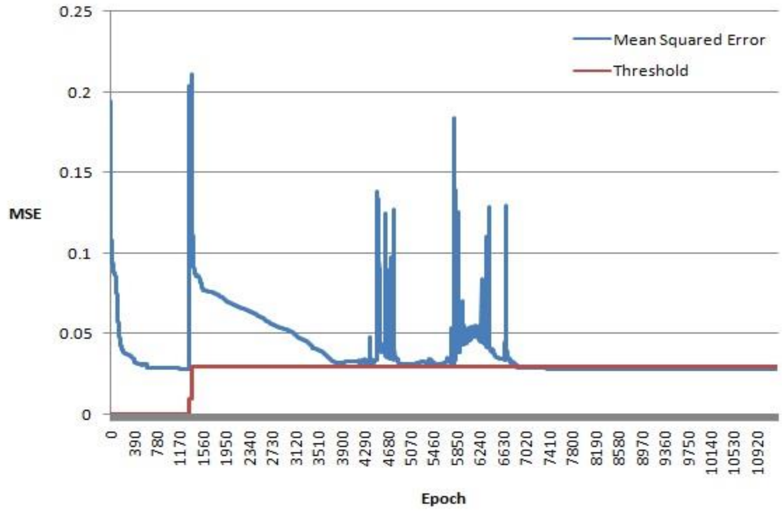

Figure 3. Convergence Curves using 40% of Iris Data Set

It has been observed in Figure 4 that the threshold remains stable until a sudden upswing in MSE triggers the gradual increase of the threshold. The sudden increase of the MSE indicates the re-initialization of the network. This re-initialization mechanism allows the network to retrain itself and verify if it has already reached its maximum potential to learn based on the presented training data set. In this experiment, the standard BP algorithm took 11,282 epochs to realize that it already reached its global minima and halt the training.

4.3. Experiment 3

The third experiment proved that the number of epoch increases as the number of training set increases and this causes the network to take a longer training time compared to the previous experiments.

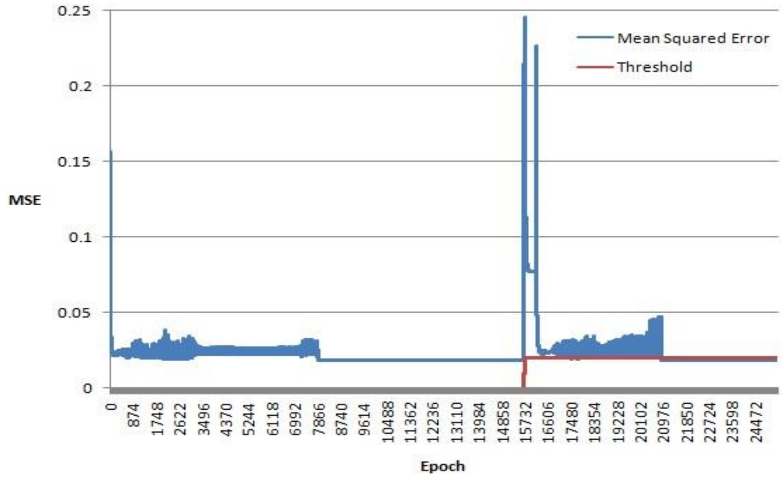

Figure 4. Convergence Curves using 60% of Iris Data Set

On contrary, the peak of the threshold in the 3rd experiment achieved the lowest error rate increasing the accuracy rate of the networks, as shown in Figure 5.

Copyright ⓒ 2014 SERSC

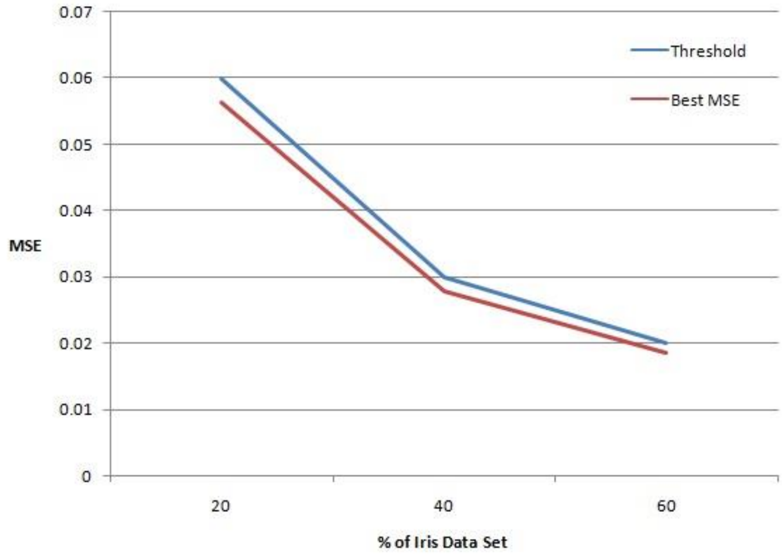

Figure 5. Distance between the Threshold and best MSE

For each experiment, the accuracy of the network, trained using the standard BP with the proposed stopping criterion, was tested using the corresponding test data sets. Figure 6 shows that the size of the training set is indirectly proportional to the distance between the threshold and best mean squared error.

Table 1. Comparison Table using Iris Data Set

% Training Set	Total Epochs	Threshold (1)	Best MSE (2)	Distance (3)	Accuracy using Test Data
20	4,192	0.06	0.05625	0.00375	92.50%
40	11,282	0.03	0.02791	0.00209	95.55%
60	25,344	0.02	0.01855	0.00145	96.66%

Table 1 clearly shows that the threshold is becoming closer to the global minima as the size of the training set increases and a notable significant increase in the accuracy rate of the network was also observed.

5. Conclusion

In this study, an adaptive stopping criterion was developed and tested using a multivariate data set, the Iris data set. The data set was divided into two different parts, the training and testing data sets, to avoid the problem of over-fitting. Three experiments were conducted to evaluate and analyze the effect of the proposed stopping criterion on the performance of the backpropagation neural networks. Based on the results of the three experiments, the proposed algorithm enabled the neural networks to learn to its maximum potential based on the patterns presented on it. Furthermore, it helped the neural networks

 Copyright ⓒ 2014 SERSC

to effectively recognize its minimum error rate, thus, allowing it to converge to the global minima.

References

[1] C.U. Joy, Comparing the Performance of Backpropagation Algorithm and Genetic Algorithms in Pattern Recognition Problems. International Journal of Computer Information Systems. vol. 2, no. 3, **(2013)**, pp. 7-12.

[2] A. Gupta and M. Shreevastava, Medical Diagnosis Using Backpropagation Algorithm. International Journal of Emerging Technology and Advanced Engineering. vol. 1, no. 1, (**2011)**, pp 55-58.

[3] L. Liang and D.Wu, An Application of Pattern Recognition on Scoring Chinese Corporations Financial Conditions Based on Backpropagation Neural Network. Computers and Operations Research. vol. 32, no. 5, **(2005)**, pp. 1115-1129.

[4] T. Sun and F. Tien, Using Backpropagation Neural Network for Face Recognition with 2D + 3D Hybrid Information. Expert Systems with Applications. vol. 35, no. 1-2, **(2008)**, pp. 361-372.

[5] E. Vamisidhar, K.V.S.R.P. Varma, P. SankaraRao and R. Satapati, Prediction of Rainfall Using Backpropagation Neural Network Model. International Journal on Computer Science and Engineering. vol. 2, no. 4, **(2010)**, pp. 1119-112.

[6] K. Burse, M. Manoria, and V.P.S. Kirar. Improved Backpropagation Algorithm to Avoid Local Minima in Multiplicative Neuron Model. Proceedings of the International Conference of AIM, **(2011)**; Nagpur, Maharashtra, India, pp 67-73.

[7] C. Chakraborty, Structural Characterization of Worm Images Using Trace Transformation and Backpropagation Neural Network. International Journal of Signal Processing, Image Processing and Pattern Recognition. vol. 5, no. 3, **(2012)**, pp. 27-48.

[8] S. Shelke and S. Apte. A Multistage Handwritten Marathi Compound Character Recognition Scheme Using Neural Networks and Wavelet Features. International Journal of Signal Processing, Image Processing and Pattern Recognition. vol. 4, no. 1, **(2011)**, pp. 81-94.

[9] J. Dheeba and G.W. Jij, Detection of Microcalsfication Clusters in Mammograms Using Neural Network. International Journal of Advanced Science and Technology. vol. 19, **(2010)**, pp. 13-22.

Authors

Jeremias T. Lalis, received his B.S. degree in Computer Science from La Salle University, Philippines in the year 2005 and M. degree in Information Technology from Cebu Institute of Technology – University of Cebu City, Philippines in 2011. He is currently pursuing his D. degree in Information Technology in Cebu Institute of Technology-University, Philippines. His research interest includes image processing, computer vision, pattern recognition, machine learning and data mining.

Bobby D. Gerado, he finished his B.S. in Electrical Engineering from Western Institute of Technology, Philippines in 1994, M.A. Ed Mathematics from the University of the Philippines, Diliman in 2000, and Ph.D. in Information and Telecommunications Engineering from Kunsan National University, South Korea in 2007. Now, he is a Faculty of Information Technology and Mathematics at West Visayas State University, La Paz, Iloilo City. His research interests lie on the area of distributed systems, data mining, ubiquitous computing, mobile communications, and Statistics.

Copyright ⓒ 2014 SERSC

Yung-Cheol Byun, he received the B.S. degree in computer engineering from Jeju National University, Korea in 1993. He received the M.S. and Ph.D. degree in computer science from Yonsei University, Korea in 1995 and 2001, respectively. He joined Electronics and Telecom- munications Research Institute (ETRI) as a senior researcher in Fall 2001. He was promoted to join the dept. of Telecommunication and Computer Engineering at Jeju National University in 2002, and he is currently a Full Professor. His research interests include USN and RFID middleware, ubiquitous computing, Location sensing using pattern recognition methods, and intelligent computing.

Copyright ⓒ 2014 SERSC