Ubiquitous Stopping Criterion for Backpropagation Learning in Multilayer Perceptron Neural Networks

J. T. Lalis[*], B. D. Gerardo[**], Y. Byun[***] and Y. Ha[****]

[*]College of Computer Studies, Cebu Institute of Technology-University, Cebu City, Philippines
[**]Institute of ICT, West Visayas State University, Luna St., Lapaz, Iloilo City, Philippines
[***]Dept. of Computer Engineering, Jeju National University, Jeju, Korea
[****]Dept. of Computer Science, Konkuk University, Seoul, Korea

j_lalis@yahoo.com, bgerardo@wvsu.edu.ph, ycb@jejunu.ac.kr and ygha@konkuk.ac.kr

Abstract. In this paper, the Fisher's Iris data set was used as input for standard BP. The network was trained using the different training set sizes and comparisons are made on the effectiveness of the proposed stopping criterion. The accuracy of the network was also tested using the testing sets. 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. The proposed algorithm is capable of identifying the minimum acceptable error rate causing the network to learn to its maximum potential based on the patterns presented on it.

1. Introduction

The structure of artificial neural networks mimics the biological neurons of the brain in recognizing patterns and learning. ANNs have been widely used in various fields due to its capability to build model for linear and non-linear systems even in the unknowingness of the relationship between its input and output data [1][2].

In 1969, Arthur E. Bryson and Yu-Chi first described Backpropagation as a multi-stage dynamic system optimization method and later gained recognition in 1986 through the work of Geoffrey E. Hinton, David E. Rumelhart, and Ronald J. Williams. This led to the revival of BP in the field of ANN research. It has been proven that multilayer perceptron (MLP) neural networks trained using the backpropagation learning method provide ubiquitous yet powerful tools for analyzing complex and real world problems such as medical diagnosis [2], financial diagnosis [3], security [4], agriculture [5] and etc.

However, standard BP algorithm suffers in some major shortcomings like slow learning speed and the problem of falling into local minima [6]. Because of it, numerous studies and modifications had already been conducted to improve it. But one

ISA 2013, ASTL Vol. 21, pp. 294 - 298, 2013

of the key elements of BP has been scarcely studied, the stopping criteria of BP. The halting of the training of both standard and modified BP depends heavily on human decision. Some studies limit the training on the number of rounds with the given patterns, known as epoch, and some are based on the predefined acceptable error rate. But the first approach limits the BP to its potential to learn more and further minimize its error rate. On the other hand, the second approach may prevent the BP to stop learning if the given acceptable error rate is beyond its learning capability based on the presented training data set.

2. Review of Related Literature

A. The Standard Backpropagation Algorithm

BP is one of the simplest and most general methods for the supervised training and data mining [5]. The standard BP works as follows:

1 Initialize all weights W and node threshold to small pseudorandom numbers.

2 Calculate the activation level O_j of the hidden and output unit:

$O_j = F(\Sigma W_{ji} O_i - \theta j)$

where W_{ji} is the weight from and input O_i, θ_j is the node threshold, and F is a sigmoid function:

$F(v) = 1/(1+e^{-v})$

3 Update the weights by starting at the output units and work backward to the hidden layers recursively:

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji}$$

where $W_{ji}(t)$ is the weight from unit i to j at the time t (or rth iteration) and ΔW_{ji} is the weight adjustment. The weight change is computed by

$\Delta W_{ji} = \eta \delta_j O_i$

where η is a trial independent learning rate ($0<\eta<1$, e.g. 0.3) and δ_j is the error gradient at unit j.

4 Compute the error gradient using the delta rule:

$E = \Sigma_j 1/2(t_j - O_j)^2$

Where t_j is the desired / target output activation and O_j is the actual output activation at output unit j.

5 Repeat step 2 to 5 until convergence on the selected error criterion. An error criterion can be either when the error is below or equal to predefined acceptable error rate or by limiting the number of epochs.

3. The ubiquitous stopping criterion algorithm

This study proposes a ubiquitous stopping criterion for BP algorithm. The algorithm is designed to dynamically adjust the stopping criterion, the *threshold*, based on the learning capability of the BP algorithm against the given training data set.

```
set threshold to 0.00
set ctr to 0
set prevMSE and leastMSE to 1
set stopTraining to False

initialize network
Round (threshold to two decimal places)
while (!stopTraining) {
  set MSE to network.Train(patterns).MSE
  increment epoch by 1
  Round (MSE to five decimal places)
  if (prevMSE is equal to MSE)
        decrement ctr by 1
      else
        increment ctr by 1
  set prevMSE to MSE
       if (MSE is less than leastMSE)
        set leastMSE to MSE
      if (ctr is equal to 0) {
            set threshold to (threshold * 3 + leastMSE) / 2
                   if (threshold is greater than leastMSE)
             set threshold to leastMSE
        if (MSE is less than threshold)
              set stopTraining to True
        else
              initialize network
  }
}
```

The rounding off of the mean squared error into five decimal places signifies minimal changes in the movement of the patterns based on the calculated delta rule.

4. Experiments and results

The Fisher's Iris data set has been used as input for the standard BP algorithm in order to evaluate the performance of the proposed stopping criterion. This data set contains 50 instances for each of the three classes. Each class refers to a certain type of Iris plant, wherein, one class is linearly separable from the other classes. The data set was divided into two parts for each of the three major rounds of experimentation: 20% for training and 80% for testing; 40% for training and 60% for testing; and 60% for training and 40% for testing, respectively.

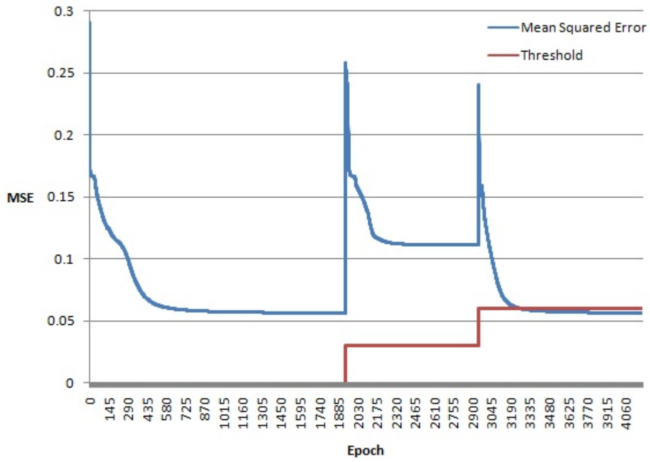

Fig. 1. Convergence curves using 20% of Iris Data Set

5. Conclusion

In this paper, the researcher proposed a ubiquitous stopping criterion which enabled the network to learn to its maximum potential based on the given patterns. The efficiency of the proposed algorithm was tested using the three different training sizes (20%, 40%, and 60% of Fishers Iris data set). The experiments results showed that the proposed stopping criterion helped the network to recognize its maximum error rate and converge to its global minimum.

Acknowledgement

This research was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (grant number: 2012006817).

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. 5, pp 7-12, 2011.
2. A. Gupta, M. Shreevastava. “Medical diagnosis using backpropagation algorithm”, International Journal of Emerging Technology and Advanced Engineering, vol. 1, no. 1, pp 55-58, 2011.

3. L. Liang, 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, pp 1115-1129, 2005.
4. T. Sun, F. Tien. “Using backpropagation neural network for face recognition with 2D + 3D hybrid information”, Expert Systems with Applications, vol. 35, no. 1-2, pp 361-372, 2008.
5. E. Vamisidhar, K.V.S.R.P. Varma, P. Sankara Rao, R. Satapati. “Prediction of rainfall using backpropagation neural network model”, International Journal on Computer Science and Engineering, vol. 2, no. 4, pp 1119-1121, 2010.
6. K. Burse, M. Manoria, V.P.S. Kirar, “Improved backpropagation algorithm to avoid local minima in multiplicative neuron model”, In Proceeding of the international conference, AIM 2011, Nagpur, Maharashtra, India, pp 67-73, 2011.