International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-02, Issue-04 E-ISSN: 2347-2693

Comparative Analysis of various Performance Functions for Training a Neural Network

Shobhit Kumar[1*], Vikash Kumar Mishra[2], Sapna Singh[4] and Neeraj Vimal[5]

[1*,4] *Information Technology, Uttar Pradesh technical University, India*

[2,3]*Computer Science and Engineering, Uttar Pradesh technical University, India*

www.ijcseonline.org

Received: 16/03/2014 Revised: 27/03/2014 Accepted: 19/04/201x4 Published: 30/04/2014

***Abstract*—** Handwriting Recognition (or HWR) is the ability of a computer to receive and interpret comprehensible handwritten input from sources such as paper documents, photographs, touch-screens and other devices. The image of the written text may be sensed "Offline" from a piece of paper by optical scanning (optical character recognition) or intelligent word recognition. Neural Network concept is the most efficient recognition tool which is dependent on sample learning. Mean square error function is the basic performance function which is most broadly used and affects the network directly. Various performance functions are being evaluated in this paper so as to get a conclusion as to which performance function would be effective for usage in the network so as to produce an efficient and effective system. The training of back propagation neural network is done with the application of Offline Handwritten Character Recognition using MATLAB simulator.

***Index Term*—** Back Propagation Algorithm, Performance Function, Mean Square Error Algorithm

I. Introduction

In a wide range of applications Backpropagation (BP) Neural Network classifiers shows good performance. Training this classifier proves to the minimization of error cost function over an accessible training set.

For pattern classification in general and for BP in particular, the cost function that is more efficiently being used than any other alternative is the standard mean square error (MSE).The standard mean square error is advantageous for having no prior knowledge for class distributions [1].This paper proposes training BP neural network using various cost error function so as to decide which performance function is beneficial for a proposed network. The MSE is optimal in yielding a network output with minimum variance for asymptotically large number of statistically independent training pattern. This paper also compares the performance of BP neural network trained by the various performance functions.

The paper is organized as follows: Section 2 introduces the various and other performance functions used in neural network. Section 3 gives an overview of Backpropagation algorithm. Section 4 describes the character recognition technique .Section 5 proposed mean square algorithm concept by training a particular input binary data image of 900 pixel value. Section 6 describes the various research methodologies which are being utilized in the research work with Section 7 giving the comparative analysis of various performance functions. Finally, the research is being concluded along with the illustrations of the results.

II. Performance metrics

Artificial Neural Network is a representation of human brain that tries to learn and simulate its training input patterns by a predefined set of example patterns. The network is trained with particular specifications. The obtained output after training the network is compared with the desired target value and error is calculated based upon these values [2].

For training an input pattern and measuring its performance, a function must be defined. The various functions being included in neural network are:

1.) Sum Of Squared error (SSE)

The first basic cost evaluation function. The Sum of Squared error is defined as

$$\mathrm{SSE}=\sum_{p=1}^{p}\sum_{i=1}^{N}(t_{pi}-y_{pi})^2$$

Where, t_{pi} = Predicted value for data point i;

y_{pi} =Actual value for the data point i;

N = Total number of data points

2.) Mean squared error (MSE)

The most widely used function and whose performance function shows efficient results. The *Mean Squared error* is defined as

$$\mathrm{MSE}=\frac{1}{N}\sum_{p=1}^{p}\sum_{i=1}^{N}(t_{pi}-y_{pi})^2$$

Where t_{pi} = Predicted value for data point i

y_{pi} =Actual value for the data point i;

N = Total number of data points

3.) Root Mean Squared error(RMSE):

The *root Mean squared error* value for finite set of data is defined as

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{p=1}^{p}\sum_{i=1}^{N}(t_{pi}-y_{pi})^2}$$

Corresponding Author: Shobhit Kumar

Information Technology, Uttar Pradesh technical University, India

Where t_{pi} = Predicted value for data point i;
y_{pi} =Actual value for the data point i;
N= Total number of data points

4.) Mean Magnitude of Relative Error (MMRE) [4,6]:

The *mean magnitude relative error* is defined as

$$\text{MMRE}=\frac{1}{N}\sum_{p=1}^{p}\sum_{i=1}^{N}\frac{|t_{pi}-y_{pi}|}{y_{pi}}$$

Where t_{pi} = Predicted value for data point i;
y_{pi} =Actual value for the data point i;
N = Total number of data points

5.) Relative Absolute Error (RAE):

The *Relative absolute error* is defined as the summation of the difference between predictive value and given value for the sample case j to that divide it by the summation of the difference between the given value and average of the given value. The relative absolute error of individual data set j is defined as

$$\text{RAE}=\frac{\sum_{i=1}^{N}|t_{ij}-y_{ij}|}{\sum_{i=1}^{N}|y_i-y_m|}$$

Where t_{ij} = Predicted value by the individual dataset j for data point in i;
y_i =Actual value for the data point i;
N = Total number of data points
y_m = Mean of all y_{pi}

6.) Root Relative Squared Error (RRSE):

The *root relative squared error* of individual data set j is defined as

$$\text{RRSE}=\sqrt{\frac{\sum_{i=1}^{N}(t_{ij}-y_{ij})2}{\sum_{i=1}^{N}(y_i-y_m)2}}$$

Where t_{ij} = Predicted value by the individual dataset j for data point in i;
y_i =Actual value for the data point i;
N = Total number of data points
y_m = Mean of all y_{pi}

7.) Mean Absolute Error (MAE)

The *mean absolute error* measures of how far the estimates are from actual values. It could be applied to any two pairs of numbers, where one set is "actual" and the other is an estimate prediction.

$$\text{E}=\frac{1}{N}\sum_{p=1}^{p}\sum_{i=1}^{N}|t_{pi}-y_{pi}|$$

Where t_{pi} = Predicted value for data point i;
y_{pi} =Actual value for the data point i;
N = Total number of data points.

The above equation represents the output nodes, t_{pi} **and** y_{pi} which are target and actual network output unit on the p^th^ pattern, respectively.

8.) Mean Square Error with Regularization (MSEREG):

The *standard Mean square error with Regularization* value can be evaluated by the formula

$$\text{MSE}=\frac{1}{N}\sum_{p=1}^{p}\sum_{i=1}^{N}\left(t_{pi}-y_{pi}\right)^2$$

$$\text{MSEREG}=\alpha*\text{MSE}+(1-\alpha)*\sum_i(W_i)^2$$

Where t_{pi} = Predicted value for data point i;
y_{pi} =Actual value for the data point i;
N = Total number of data points
α = Performance Ratio
W_i= Weight of the network

The network learns by adjusting weights. The process of adjusting the weights to make the neural network learn the relationship between the input and targets is known as learning or training. There are several techniques for training a network gradient descent method which is the most common.

III. THE BACKPROPAGATION APPROACH

The backpropagation neural network is a neural network with a layered, feed-forward network structure and the generalised delta rule which updates its weight for each run of the network [5]. It is a powerful mapping network which has been successfully applied to a wide variety of problems. Backpropagation is one of the finest algorithms that give better performance with relatively more accurate recognition ratio. Training is proficient by presenting the patterns to be classified to the network and determining its output. The actual output of the network is compared with the target output and an error measure is calculated. The higher the error value the less efficient the network is. Thus an effort has been made to decrease the error value for the Backpropagation network.

IV. CHARACTER RECOGNITION

A generic character recognition (CR) system is a sequence of steps. Its different stages are as given below:

1. **Input**: Samples are read to the system by a optical scanner.
2. **Preprocessing**: Pre-processing converts the image into a digital form suitable for subsequent processing and feature extraction [3].
3. **Segmentation**: The most basic step in CR is to segment the input image into individual glyphs. This step separates out sentences from text and subsequently words and letters from sentences.
4. **Feature extraction**: Extraction of features of a character forms a vital part of the recognition process [4]. Feature extraction captures the vital details of a character.
5. **Classification**: During classification, a character is placed in the appropriate class to which it belongs [6].
6. **Post Processing**: Combining the CR techniques either in parallel or in series.

V. MEAN SQUARE ERROR ALGORITHM

In statistics, the **mean squared error (MSE)** of an estimator is one of many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated. MSE is a risk function, corresponding to the expected value of the **squared error loss** or **quadratic loss**. MSE measures the average of the squares of the "errors." The error is the amount by which the value implied by the estimator differs from the quantity to be estimated. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimated value [wiki].

Minimizing squared error would increase the accuracy of a particular system with defined number of input training dataset [2]. The algorithm or the sequence of steps for the error calculation is as follows:

MSE Algorithm:

```
Initialize 'n' number of input patterns (integer values)
Do
    For each training pattern 'n' train the network
    O=neural_net_output (network,n)
    T=neural_net_target (desired)
    Compute error e=(T-O) at the output units
    Square the error 'e'
    Calculate the summation of error 'e' for all
    input patterns
While (n!=0)
   Divide the summation obtained by the 'n'
   number of patterns
```

The standard mean square error is advantageous for having no prior knowledge for class distributions. It is widely used as it results in the least error value as compared to other error values. Likewise various error functions can be evaluated along with its particular algorithm.

Minimizing error value will result in more accurate system for pattern recognition and other applications. It is advantageous for a system with large dataset, where thousands of values are to be enumerated to result thousand of values. It has proved useful in calculation of error for Backpropagation NN which back propagates its error value with the adjustment of weights for its every iteration [9].

The proposed Mean square error algorithm is a description of the way how the various error values can be enumerated and modified so as to improve its performance with reduced cost error. By enumerating the various error functions mathematically it can be drawn to a conclusion, as to when a particular performance function should be used and which of them gives reduced error cost.

VI. RESEARCH METHODOLOGY

The Artificial Neural Network (ANN) and Gradient Descent learning techniques have been used for predicting the software effort using dataset of software projects in order to compare the performance results obtained from these models.

(A) Empirical Data Collection: The data we have used is being evaluated from a binary image of 900 pixel values. The error value of a network is the mathematical value of target value difference the obtained value after training the network. The MSE is calculated for 900 pixel values. Then the other error functions are being calculated. A threshold value is set (let 0.5) so as to attain the number of training cases recognized. If the obtained output is more than the threshold value then the neural network has recognized the training pattern.

A feed-forward Backpropagation network is established using 900 input neurons, 12 hidden layer neurons and 2 output neurons. The output set is either 0 or 1.The 0 represent the pattern as unrecognized and 1 as recognized. Gradient descent with momentum algorithm with the various performance functions are being evaluation algorithm.

(B)Diagrammatical representation of error calculation: Firstly initialize the training patterns and set a particular target value for the input training pattern. Set the maximum number of iterations up to which the input data's are to be iterated. Then train the input pattern by using Back propagation training algorithm [8]. Compare the obtained output with the target value set. If it is near the target value then the network has realized the training pattern else it has not. The flow of data in the whole recognition procedure is being shown in fig 2.

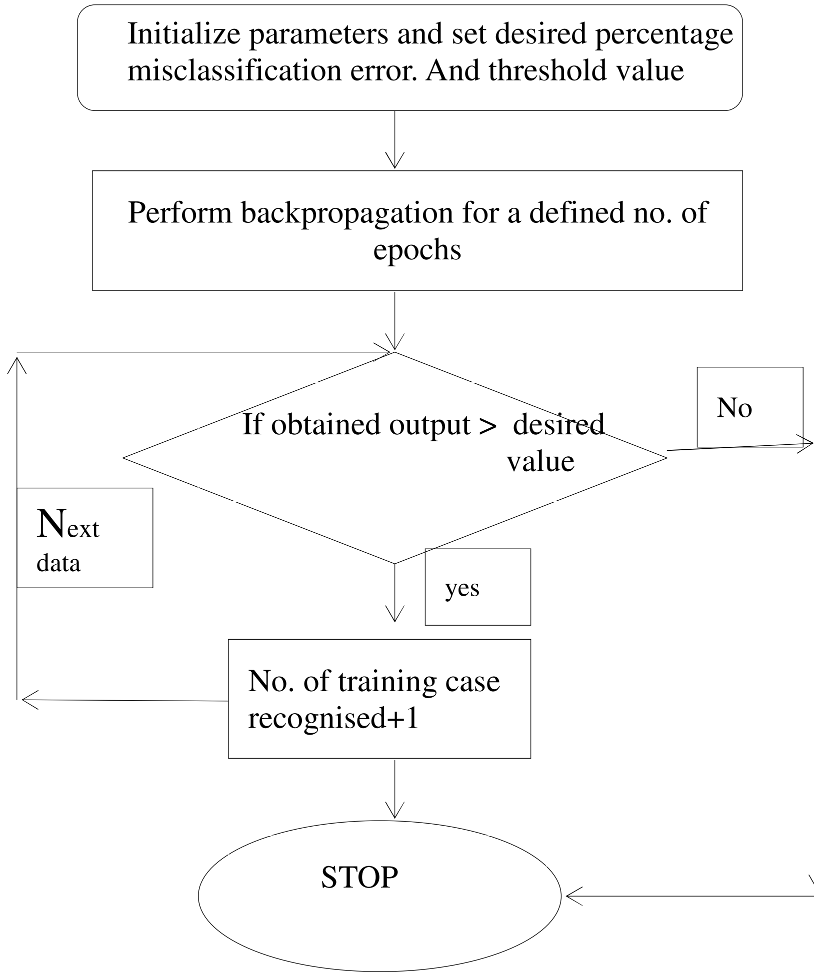

Figure2: Diagrammatical representation of number of recognition patterns

VII. PERFORMANCE COMPARISON

The various error functions are being calculated for the training input patterns .Then an estimate of which would be the efficient function for a neural network. The calculated error value is being shown through the table

	MSE	SSE	MSEREG	MMRE	RMSE	MAE
Error	0.00096	0.00162	0.00086	0.0244	0.0308	0.0561

Table 1: Comparative analysis of various Error functions

The readings are being obtained by training the Backpropagation neural network by the various error functions. As being concluded from the readings above that the *mean square error with Regularization* functions shows positive results with minimum error as compared to the other evaluated functions. The result is obtained from 100 character values through binarization of input data. The lesser the value of error signifies the more efficient and cost reduced the network is being devised.

VIII. RESULT

To show that the MSEREG (Mean square error with Regularization value) formula shows better results, all the evaluated functions were being calculated for their respective MSE. The tabulated data's above proves that the formula with the MSEREG (Mean square error with Regularization value) has enhanced the performance of the network.

The graph below clearly depicts the mean square error value evaluated by using various functions. The different functions are being evaluated so as to check the affinity of our proposed MSEREG function. Thus it can be said that MSEREG performance function is the most efficient and cost reduced function.

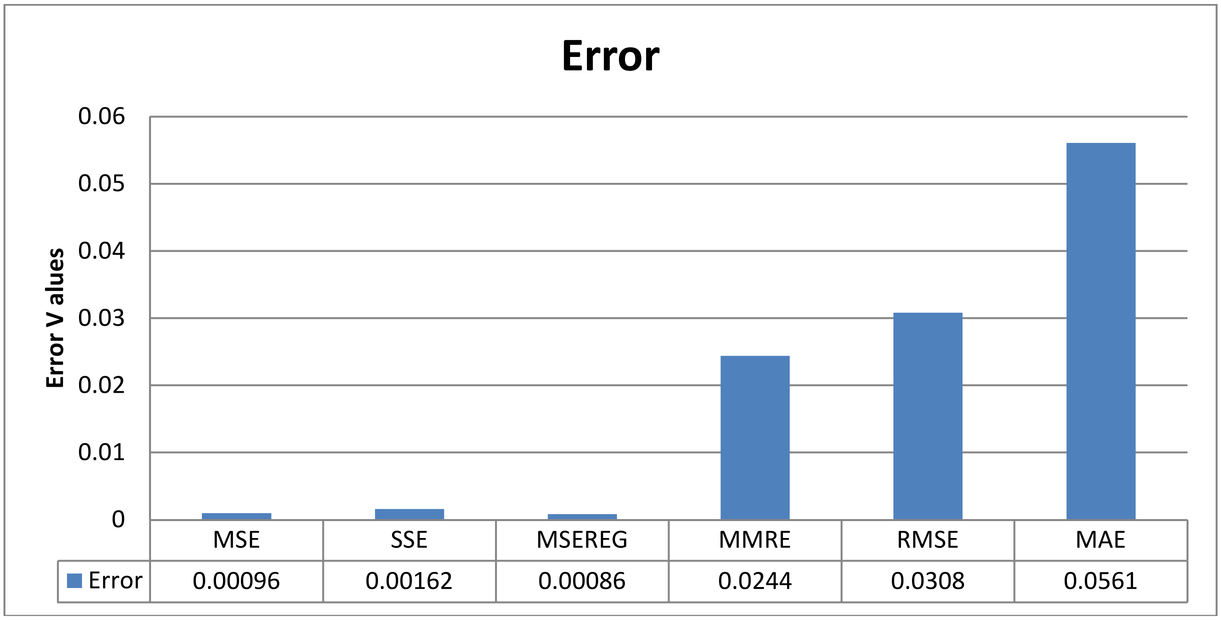

Figure3. Comparative analysis of various error functions

IX. CONCLUSION

The Backpropagation training algorithm has its advantage over training patterns as it increases the accuracy of the simulated data [10]. The mean square error Regularization value is beneficial for having no prior knowledge for class distributions. Through this paper it is being proposed that by reducing the error value, accuracy of a particular network can be increased. Further work is under process for reducing the training time and increasing the accuracy for a network using Backpropagation training algorithm with its application on hand written character recognition. This concept can be further applied to various applications related to classification and pattern recognition where larger dataset is being processed. Work can also be done to improve the accuracy and increasing recognition speed as its area of application is wide.

REFERENCES

[1] Bogdan M. Wilamowski, Serdar Iplikci, Okyay Kaynak, M. Önder Efe , "An Algorithm for Fast Convergence in Training Neural", 0-7803-7044-9/01/$10.00 © IEEE , 2001.

[2] Hussein Rady, Reyni's, "Entropy and Mean Square Error for Improving the Convergence of Multilayer Backpropagation Neural Networks: A Comparative, Study", 117905-8282 IJECS-IJENS, Vol: 11, No: 05, October-2005.

[3] Hossam Osman, Steven D. Blostan, "New cost function for Backpropagation neural network with application to SAR imagery classification", K71396.

[4] G. R. Finnie and G.E. Wittig, "A Comparison of Software Effort Estimation Techniques: Using Function Points with Neural Networks, Case Based Reasoning and Regression Models," Journal of Systems and Software, vol.39, pp.281-289, 1997.

[5] G. R. Finnie and G.E. Wittig, "AI Tools for Software Development Effort Estimation," Proceedings of the International Conference on Software Engineering: Education and Practice , SEPT- 1996.

[6] K. Srinivasan and D. Fisher, "Machine Learning Approaches to Estimating Software Development Effort," IEEE Transactions on Software Engineering, vol.21, Feb-1995.

[7] Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, L. Malik , M. Kundu and D. K. Basu, " Performance Comparison of SVM and ANN for Handwritten Devnagari Character Recognition" , IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, May-2010.

[8] Srinivasa kumar devireddy, Settipalli appa rao, "Hand written character recognition using backpropagation network", Journal of Theoretical and Applied Information Technology(JATIT) , 2005 - 2009.

[9] Madhu Shahi, Dr. Anil K Ahlawat, Mr. B.N Pandey , "Literature Survey on Offline Recognition of Handwritten Hindi Curve Script Using ANN Approach" ,International Journal of Scientific and Research Publications, Volume 2, Issue 5,ISSN 2250-3153, May-2012.

[10] Naveen Garg, Sandeep Kaur, "Improvement in Efficiency of Recognition of Handwritten Gurumukhi Script" ,IJCST Vol. 2, Issue 3, September-2011.

[11] Rakesh Kumar Mandal and N R Manna, "Handwritten English Character Recognition using Pixel Density Gradient Method", International Journal of Computer Sciences and Engineering, Volume-02, Issue-03, Page No (1-8), Mar -2014.

