Indian Journal of Science and Technology, Vol 8(16), DOI: 10.17485/ijst/2015/v8i16/62540, July 2015
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

DI-ANN Clustering Algorithm for Pruning in MLP Neural Network

P. Monika* and D. Venkatesan

School of Computing, SASTRA University, Thanjavur - 613401, Tamil Nadu, India;
monikapugalendhi@gmail.com

Abstract

Data mining is an emerging technology for applications such as text based mining, web based mining and it performs a major role in various domains for numerical data analysis, data statistics and machine learning. In this paper, data mining is used in machine learning of ANN (Artificial Neural Network). The Pruning technique in an MLP (The Multilayer Perceptron) neural network is to remove the unwanted neurons based upon their corresponding weights; as a result it improves the accuracy and speed of the network. In the existing system, based on their synaptic weights the pruning is performed and it removes the lowest weight neuron form the network. The result obtained from the existing method does not produce an optimized removal of the neuron. In the proposed system pruning is performed by using divisive clustering in MLP neural network. The main purpose of the Divisive algorithm in ANN is to split each neuron weight into sub neuron up to the fixed level and then remove the least weighted hidden neuron. The proposed method is implemented using the Java language. The Performance result obtained from the proposed method shows that it reduces the error rate and improves efficiency and accuracy of the MLP network. The present results confirm that DI-ANN (Divisive Artificial Neural Network) can provide a fast, accurate, and consistent methodology applicable to the neural network.

Keywords: DI-ANN (Divisive Artificial Neural Network) Algorithm, MLP (Multi Layer Perceptron), Neural Network (NN), Pruning Method

1. Introduction

Data mining technology is used in neural network to extract the hidden information from the large data set. A classification problem in data mining is performed based upon the probability of data and then it is grouped into particular classes. While classifying the data set it involves two steps, the first step is learning and the second step is to classify the test data accurately based on classification rules.

The Machine Learning (ML)[12] algorithm uses the data mining technique to classify the different data sets. The main purpose of Machine Learning algorithms is used to train the large collection of patterns. Once training is performed, then it starts to test the new patterns which are not used during the training, in order to check how the system is able to find the pattern beyond the training data.

The Multilayer Perceptron(MLP)[13] is one of the feed forward ANN (Artificial Neural Network). It consists of three or more layers (input, output and hidden layers) each node in the layer is associated with some weights Wij. MLP network is a directed graph in which all neurons of one layer are fully connected to all other neurons of the next layer. This network uses the supervised learning approach when it has a set of training data.

The supervised learning algorithm contains the desired output of all the training examples. Back Propagation algorithm[14,15] is used to train the data set to predict the error. The error obtained from this algorithm is the difference between actual and desired output. When a neural

*Author for correspondence

network size is big it takes more time to training the input values, so the pruning algorithm is used to trim the size of the network. As a result of applying the pruning algorithm the network size becomes reduced and it avoids the unwanted complexity and over fitting problem that usually occurs in machine learning.

Divisive clustering in Artificial Neural Networks (DI-ANNs) is used to prune the neural network. This divisive algorithm is a top down approach in which all the weights obtained from each layer is formed into a single cluster. It is split by using the flat clustering algorithm; this procedure continues recursively to each data's. Divisive algorithm in ANN applies to the input parameters; it splits each neuron weight into the sub neuron to prune the least weighted hidden neurons. By pruning the network it becomes fast, accurate and consistent.

2. Literature Survey

Sharma UM[1] proposed a method for the development of face recognition and verification system by using the ANN (Artificial Neural Network) and Principal Component Analysis method. In this paper Back propagation algorithm is used to improve the performance and it also tolerates various noises based upon the recognition rates.

Gharehchopogh FS, Khaze SR, Maleki I[2] proposed an ANN (Artificial Neural Network) and KNN algorithm which is used to classify the bloggers, in web logs the classification is one of the important criteria. The performance shows that it has more improvement when compare to earlier approaches and it provides good results for bloggers to classify.

Molani M, Ghaffari A, Jafarian A[3] proposed an algorithm to predict software project cost and it is estimated by using the Artificial Neural Network (ANN). To estimate the project cost usually it needs a complex algorithm to calculate cost, by using neural network it eliminates the complex algorithm. Subramanian JV, Sadiq MAK[4] proposed a Neural Network algorithm to predict the location of the user using mobile. User Pattern learning approach is used and it updates the signal cost based upon the location. The performance result shows the efficiency and effectiveness of the pattern learning approach.

OPBDM (Optimized Prune Based Data Mining) has been proposed[5] mainly in making decisions for the given data set and also produces the decision support model for the data. It performs post pruning in spatial databases to construct the tree for various data sets. This OPBDM reduces the complexity, redundancy and noisy data in databases are removed. The problem in OPBDM is that it does not explain about the partial information hiding, such as replacement of redundancy.

Le Cun Y, Denker JS, Solla SA[6] proposed an Optimal Brain Damage (OBD) method which is used to prune the elements based upon their weights. It removes the excessive weight over the training data so that it reduces the size of the network without affecting the validation. It reduces the complexity as a result, it improves the speed of learning/classification, and better generalization and it require only less training examples. These methods have some limitation in which it has the low computational efficiency and it might remove the wrong weights.

Miriam Rodrigues Silvestre, Lee Luan Ling[7] proposed the APERTP (Apparent Training Proportional Error Rate) method to prune the network. It performs pruning by removing the lowest synaptic weight value on each step. The disadvantage of APERTP is that it contains more number of lowest weight neurons are available in the network, it does not produce optimized network.

Yang J, Bouzerdoum A. and Phung S[8] proposed Compressive Sampling (CS) based pruning techniques. This method locates the significant element without calculating the saliency and it removes the least weight and hidden neurons from the network. As a result, it constructs a better topology; improve the accuracy and computational complexity. The main drawback of this method is it takes more time to train the network.

Thuan Q. Huynh, Rudy Setiono[9] proposed a pruning method called Cross validation, in which the data set is divided into two: one is cross validation set and other one is training data set. The pruning criteria are based upon the magnitude value of each weight and the cross validation step is added and it is used to test the prune network. There is a problem with this technique is pruning is based on the weight magnitude, so there is a chance to remove the important parts of the network.

3. Proposed Methodology

In this paper, proposed method uses the divisive clustering is to split the each neuron weight into sub neurons to remove the least weighted neurons from the network. The cluster is done by considering all the weights and then form into groups. The steps involved in the proposed

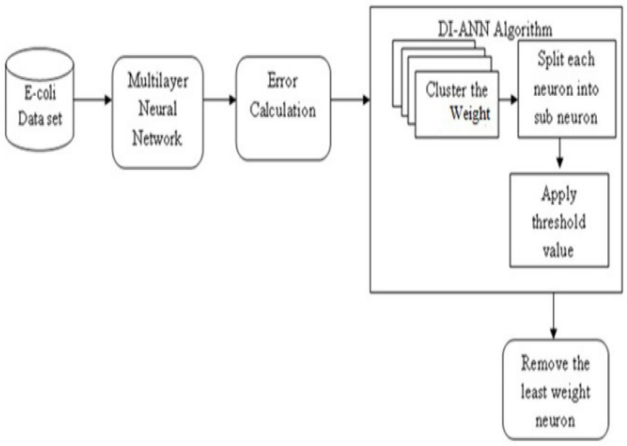

Figure 1. System Architecture.

method are given below, and overall method is represented in Figure 1.

3.1 Data Set Description

The attributes of the data set used in this paper are described in the following Table. The data set descriptions are available on UCI web Machine Learning repository[11]. The main purpose of this data set is to predict the protein Localization sites in *E. coli* bacteria. The *E. coli* data sets are given as an input to the neural network to predict the correct output.

Table 1. *E.Coli* data set

Data set	No. of instances	Input attributes	Output Value	Total attributes
E.coli	336	7 (predictive)	1 (name)	8

3.2 Neural Network Initialization

The MLP neural network is one of the most commonly used feed forward technique for classification algorithms contains large data's. The neural network is initiated by taking the *E. coli* data set as an input. It consists of totally three layers: input layer, hidden layer and output layer and each neuron are associated with weights. The formula for MLP neural network to predict output is:

$$O_j = \phi\left(\sum_{i=1}^{n}[w_i.in_i + b) \ldots\ldots\ldots\ldots(1)\right]$$

Where O_j is an output, ϕ is an activation function, w_i is weight for each input, ini is an input value and b is bias.

The training of the data set is performed by supervised learning, these training data contains the input values and it is connected to the correct output values. The In1, In2……In7 are the seven input values, out1 is the corresponding output and layer1 is the input layer and layer2 are hidden layer as shown in the Figure 2.

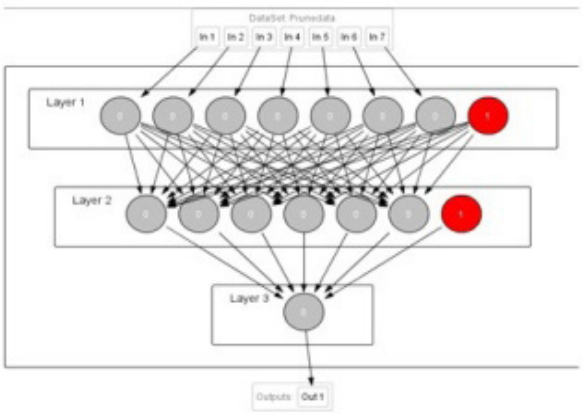

Figure 2. Network Initialization.

The output values are termed as target values. The training is carried out by the back propagation algorithm and it is used to find the error value and mean square error of the neural network.

$$error_i = targetOut_i - calOut_i \ldots\ldots\ldots(2)$$

In the above formula, $calOut_i$ is the predicted output value, $targetOut_i$ is desired output and errori represent the difference between desired and predicted value. The MSE (Mean Squared Error) is the total average of the network error with different training examples and it is also defined as the square of the average error value. When the network has least MSE then it produces an optimal neural network.

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(error_i)^2 \ldots\ldots\ldots\ldots(3)$$

Where n is the no of total neurons in the network.

3.3 DI-ANN (Divisive Artificial Neural Network) Algorithm

The Divisive algorithm[10] is the one of the hierarchical clustering technique and it is a top-down approach. This

clustering is used in ANN (Artificial Neural Network) to prune the data sets; it is applied to all weights obtained from each layer.

The clustering start with all the weights in one cluster and then dissimilarity average is calculated for each data present in the cluster. The weight which having the highest dissimilarity value is added to the next new cluster, this process repeat until no dissimilarity is available between the clusters. The maximum distance between the data's in cluster is calculated by using the formula,

$$dis(a,b) = (a_1 - b_1)^2 + (a_2 - b_2)^2 \ldots + (a_n - b_n)^2 \ldots\ldots (4)$$

In the above formula where dis (a,b) is the distance between weight values present in each cluster.a_1, b_1, a_2, b_2……a_n, b_n are the weights. The proposed algorithm logical steps can be represented in Figure 3.

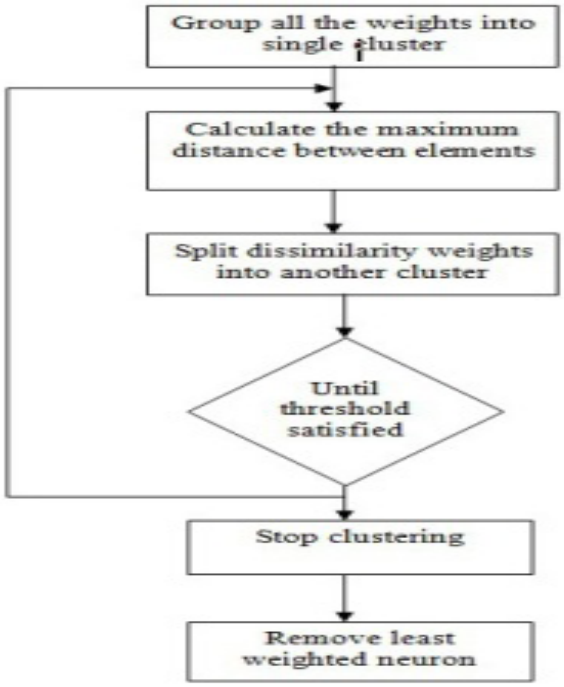

Figure 3. DI_ANN Algorithm.

3.4 Remove Unwanted Neuron

After clustering is performed, all the weights which are having the least weighted hidden neurons are removed from the network. Thus pruning is performed by removing least weight from the network; as a result the network size is trimmed (reduce). The neurons are removed and again it performs the training and it calculates the error rate to check the efficiency and accuracy of the neural network.

3.5 Pseudo Code for Proposed Algorithm

The DI-ANN algorithm is used as a proposed method for pruning

Step 1: Start with all the weights in one cluster.

Step 2: Repeat the step recursively until all the clusters become singleton

- Choose all the weights in one cluster C_1
- Remove the weight in C1 with largest dissimilarity compared with all other weights in C, and add that weight is added into new cluster C_2.
- The dissimilarity between cluster is calculated by using this formula,

$$diam\,(C) = \max_{a,b \in C} dis(a,b)$$

Step 3: Repeat the following step until no data in C_1 is closer to C_2.

Where C_1, C_2 are two clusters, a and b are two points of weight.

Remove the weights which having least weight as a result it reduces the hidden neuron which is having lowest weights.

4. Performance Measure

4.1 Experimental Setup

The proposed system is implemented by using the JAVA. The *E. coli* data set (7 predictive and one name) is chosen from the Table 1 and these parameters are used as an input for the MLP neural network and training is performed by using the back propagation algorithm.

Figure 4 explains the MSE (Mean Square Error rate) value obtained from the back propagation algorithm in MLP Neural Network. The MSE value is defined as the square of the difference between the target output and predicted output value. The Figure shows the comparison of the MSE value obtained from the before pruning and after pruning.

Table 2 explains the accuracy of the proposed method where it removes the no of hidden neurons after performing the pruning. The Table 2 contains the no of input values, output values and hidden neurons. The removal of hidden neuron is performed by using the formula,

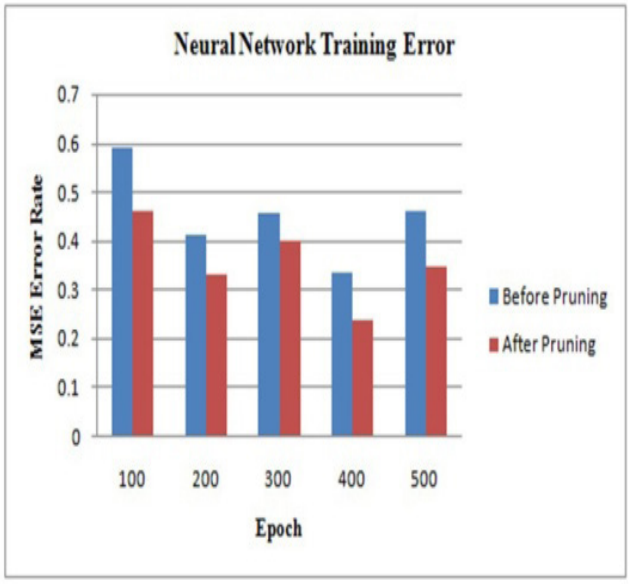

Figure 4. Comparison of Error Rate on before and after pruning.

Table 2. Accuracy of Proposed method

Description	Input values	Output value	No of hidden neurons
Before Pruning	7	1	7
After Pruning	7	1	4

$$hidden\ neurons = {}^{2}/_{3}[input\ values + output\ values] \ldots\ldots\ldots (5)$$

The hidden neurons are reduced from 7 to 4 based upon their corresponding weights.

5. Conclusion

In this paper, we proposed DI-ANN algorithm to prune the least weighted neurons, the proposed method is applied to the E.coli classification problem which contains 8 instances. The performance results show the removal of least weighted hidden neurons. DI-ANN algorithm can produce an accurate, fast and efficient method that applies to all the MLP neural networks. It also reduces the error rate and size of the network.

6. References

1. Sharma UM. Hybrid feature based face verification and recognition system using principal component analysis and artificial neural network. Indian Journal of Science and Technology. 2015; 8(S1):115–20.
2. Gharehchopogh FS, Khaze SR, Maleki I. A new approach in bloggers classification with hybrid of k-nearest neighbor and artificial neural network algorithms. Indian Journal of Science and Technology. 2015; 8(3):237–46.
3. Molani M, Ghaffari A, Jafarian A. A new approach to software project cost estimation using a hybrid model of radial basis function neural network and genetic algorithm. Indian Journal of Science and Technology. 2014; 7(6):838–43.
4. Subramanian JV, Sadiq MAK. Implementation of artificial neural network for mobile movement prediction. Indian Journal of Science and Technology. 2014; 7(6):858–63.
5. Parthasarathy G, Tomar D. Optimized Prune Based Data Mining (OPBDM) for distributed databases: an adaptive approach. Journal of Theoretical and Applied Information Technology. 2014; 67(3):717–24.
6. LeCun Y, Denker JS, Solla SA, Howard RE, Jackel LD, editors. Optimal brain damage. NIPs. 1989; 2(3):1-8.
7. Silvestre MR, Ling LL. Pruning methods to MLP neural networks considering proportional apparent error rate for classification problems with unbalanced data. Measurement. 2014; 56:88–94.
8. Yang J, Bouzerdoum A, Phung SL, editors. A neural network pruning approach based on compressive sampling. 2009 International Joint Conference on Neural Networks (IJCNN); 2009. p. 3428–35.
9. Huynh TQ, Setiono R, editors. Effective neural network pruning using cross-validation. Neural Networks. Proceedings of IEEE International Joint Conference on (IJCNN'05). 2005; 2: 972–7.
10. Savaresi SM, Boley DL. A comparative analysis on the bisecting K-means and the PDDP clustering algorithms. Intelligent Data Analysis. 2004; 8(4):345–62.
11. Machine Learning Repository. 2015 Jan. Available form: https://archive.ics.uci.edu/ml/machine-learning Databases/e.coli
12. Mitchell TM. Machine learning and data mining. Communications of the ACM. 1999; 42(11):30–6.
13. Urolagin S. Multilayer feed-forward neural network integrated with dynamic learning algorithm by pruning of nodes and connections. International Journal of Computer Applications. 2012; 47(2):7–17
14. Riedmiller M. Advanced supervised learning in Multi-Layer perceptrons-from backpropagation to adaptive

learning algorithms. Computer Standards and Interfaces. 1994; 16(3):265–78.
15. Subavathi SJ, Kathirvalavakumar T. Adaptive modified backpropagation algorithm based on differential errors. International Journal of Computer Science, Engineering and Applications (IJCSEA). 2011; 1(5):21–4.