IMPROVE AN EFFICIENCY OF FEEDFORWARD MULTILAYER PERCEPTRONS BY SERIAL TRAINING

[1]SHOUKAT ULLAH, [2]ZAKIA HUSSAIN

[1]Lecturer Computer Science, Govt. Post Graduate College Bannu, N-W.F.P Pakistan 28100

[2]Assistant Programmer, Elementary & Secondary Education Department Bannu, N-W.F.P Pakistan 28100

E-mail: shoukatullah@teacher.com, zakiahussain_bn@yahoo.com

ABSTRACT

The Feedforward Multilayer Perceptrons network is a widely used model in Artificial Neural Network using the backpropagation algorithm for real world data. There are two common ways to construct Feedforward Multilayer Perceptrons network, that is, either taking a large network and then pruning away the irrelevant nodes or starting from a small network and then adding new relevant nodes. An Artificial Neural Network model is often avoided due to the large size of network and the training that would be too slow to be tolerable. For improving the efficiency and to provide accurate results on the basis of same behaviour data, a serial algorithm for the training of data is proposed that uses two data mining techniques, that is, cluster analysis, which partitions large dataset into similar n blocks and then these n blocks are inputted to Feedforward Multilayer Perceptrons network to perform serial training for improving the efficiency.

Keywords: *Artificial Neural Network (ANN), Cluster Analysis, Data Mining, Data Classification*

1. INTRODUCTION

In the past several years, information technology and databases have created lots of innovations in the areas of life. More businesses and organizations are collecting high quality data on a large scale. The huge amount of data can be a gold mine for business management. It is therefore increasingly important to analyze the data. However, timely and accurately processing tremendous data analysis in traditional methods is a difficult task. The ability to analyze and utilize massive data lags far behind the capability of gathering and storing it. This explosive growth of data has generated an urgent need for new techniques that transform the vast amounts of data into useful information and knowledge.

Data mining is the analysis of (often large) observational data sets to find unexpected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner [1].

Data mining is defined as the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [2].

Data mining can be classified either as descriptive or predictive mining, which performs the tasks of classification and regression, or clustering and association respectively. Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects [3], whereas data classification is a two-step process to construct a model and then to use the model for classifying future or unknown objects [4].

In this study we realized the cluster analysis and data classification techniques. By cluster analysis technique a larger dataset is partitioned into similar n clusters and the training of clusters is carried out by feedforward Multilayer Perceptrons (MLPs), which is a widely used model in Artificial Neural Network (ANN) using the backpropagation algorithm [3] to perform blockwise or parallel training for improving the efficiency of MLPs network.

The field of neural network (NN) was originally kindled by psychologists and neurobiologists who sought to develop and test computational analogues of neurons which are building blocks of human brain and nervous system [5]. An ANN consists of an input layer, one or more hidden layers, and an

Vol6. No1. (pp 017 - 020)

output layer, where each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class of the input tuples/samples [3].

The rest of the paper is organized as follow:

Section 2 provides a review of related literature, section 3 and 4 encompass proposed architecture and algorithm respectively, section 5 includes experiments on the real dataset using Neuralworks Predict Ver 3.22, and paper is concluded in section 6.

2. RELATED LITERATURE

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects [6]:

i. Knowledge is acquired by the network through a learning process.

ii. Interneuron connection strengths known as synaptic weights are used to store the knowledge.

Artificial neural systems, or neural networks, are physical cellular systems which can acquire, store, and utilize experiential knowledge. [7]

There are many types of ANNs that has been employed and these include MLPs, radial basis function (RBF) and Bayesian neural networks [8, 9]. New pattern recognition methods called Support Vector Machines (SVMs) and Gaussian Mixture Models (GMMs) have been proposed and found to be particularly suited to classification problems [10]. The most well known types of ANN are following:

(a) *Feedforward ANN*

In a feedforward ANN, the connections between units do not form cycles. It usually produces a response to an input quickly. The Feedforward MLPs [9, 11] network is a widely used model in ANN using the backpropagation algorithm [12] for real world data. However, it is often avoided due to the large size of network and the training that would be too slow to be tolerable.

(b) *Feedback ANN*

In a feedback or recurrent ANN, there are cycles in the connections. In some feedback NNs, each time an input is presented, the NN must iterate for a potentially long time before it produces a response. Feedback NNs are usually more difficult to train than feedforward NNs.

There are a great many potential applications of neural networks. Their ability to learn from experience makes them suitable for solving a wide range of complex problems. Many of the problems they are used for have previously been hard to solve using computers.

Recently researchers are using the Word Wide Web (WWW) as an additional resource of training data for language modeling adaptation procedures [13]. Different information retrieval techniques have been used for dynamic adaptation of vocabulary and/or language model to the topics present in broadcast news using relevant documents obtained from a large general corpus or from the Web [14, 15, 16].

3. PROPOSED ARCHITECTURE

In this research paper, we proposed new architecture that combines two data mining techniques, that is, cluster analysis and data classification in order to improve an efficiency of feedforward MLPs (see Fig.1). This architecture is divided into two phases. In first phase, cluster analysis technique is applied to generate n clusters, and in the second phase, the generated clusters are fed into MLPs networks to carryout serial training to improve the efficiency.

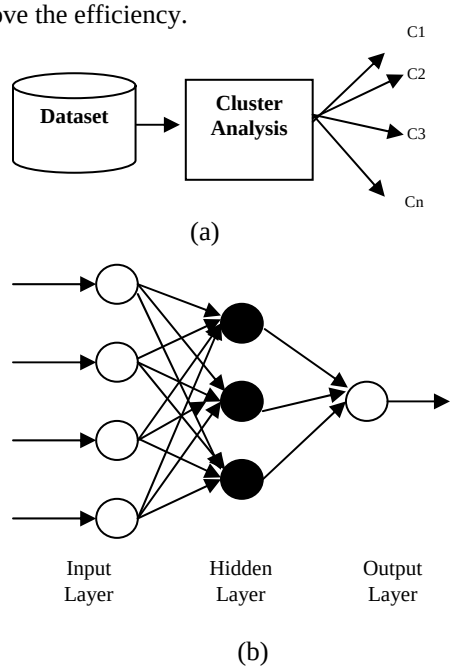

Fig. 1. (a) Cluster Analysis technique

(b) Multilayer Perceptrons network

4. PROPOSED ALGORITHM

The proposed Serial Training Algorithm (STA) combines two data mining techniques (see Fig.2). Cluster analysis technique applied in steps 1 and 2 to generate clusters C1, C2,…,Cn and steps 3 to 9 backpropagation algorithm applied which accepts n clusters as input and gives a trained neural network as output.

ALGORITHM: **STA**
//Serial Training Algorithm Based on Cluster Analysis and Backpropagation
Input: Dataset
Output: A trained neural network
(1) for each record R in a dataset {
(2) Apply Clustering Technique, and
Obtain clusters C1, C2,…,Cn }
(3) Initialize w_i and b_i // weight w & bias b
(4) Repeat Steps 5 to 10 while terminating condition is not satisfied {
(5) for each training Cluster C_j {
(6) Calculate Net input I_j and output O_j of each unit in hidden and output layers
(7) Calculate Error of each unit in hidden and output layers
(8) if Error exists Then
(9) Calculate updated weights and biases
}}

Fig. 2 The proposed algorithm

5. EXPERIMENTS

Experiments have been performed on real dataset of student's data of admission in MS-Computer Science [17] that contains one hundred and eleven (111) instances and twelve (12) attributes with a class label, in which only relevant attributes that contribute in the training process are selected by attribute selection method. The experiments are carried out in two ways using software Neuralworks Predict Ver 3.22. (see Fig. 3, 4 and 5)

(a) *Whole dataset training*

First, training on the whole dataset is performed.

(b) *Serial training*

Second, dataset is partitioned in two and three clusters and then serial training is performed. The results are presented in Table II.

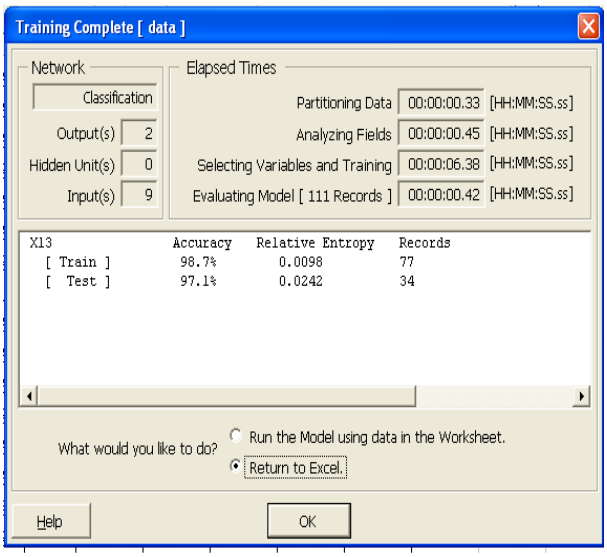

Fig. 3. Training results of whole dataset

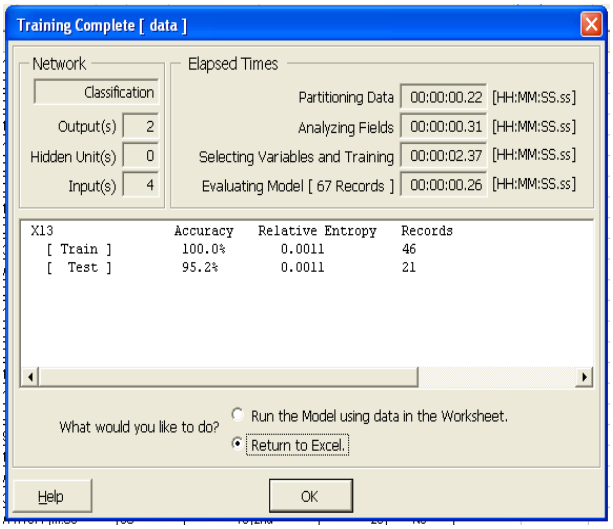

Fig. 4. Training results of cluster I

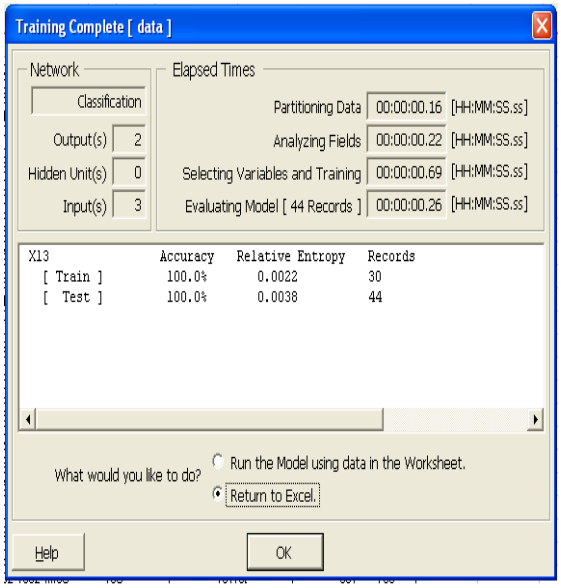

Fig. 5. Training results of cluster II

TABLE II

Results of serial training

No. of Clusters	Partition Time	Training Time(T.T)			Total Time
		C1	C2	C3	
1		6.38			6.38
2	0.30	2.37	0.69		1.83
3	0.30	0.79	1.13	0.12	0.98

6. CONCLUSION

In this research paper, we have introduced a new serial training algorithm which was able to train the real dataset in relatively short amounts of time as compared with traditional MLPs based on backpropagation algorithm. The serial training time of for two clusters is 71% faster than traditional MLPs network and for three clusters, it is 84%. Finally it is also concluded that training time decreases as number of clusters increase.

REFERENCES:

[1] H. David, M. Heikki, and S. Padhraic, "Principles of Data Mining", MIT Press, 2001Cambridge, MA.

[2]. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, "Data Mining to Knowledge Discovery: An overview", 1996, CA: AAAI Press.

[3] J. Han, and M. Kamber, "Data Mining: Concepts and Techniques", Elsevier India Pvt Limited, Second Edition, 2007, pp. 388.

[4] A. Pieter, and Z. Dolf, "Data Mining", Dorling Kindersley Publishing Inc (India).

[5] S. Richard, "Clinical Neuroanatomy", 6th Edition, Lippincot William and Walkins 530 Walnut Street Philadelphia, PA 19106.

[6] S. Haykin, "Neural Networks: A Comprehensive Foundation", NY: Macmillan, 1994, p. 2

[7] J. M. Zurada, "Introduction To Artificial Neural Systems", Boston: PWS Publishing Company, (1992), p. xv.

[8] T. Marwala, "Fault Identification Using Neural Networks and Vibration Data", University of Cambridge Ph.D. Thesis, , 2001, Cambridge, UK.

[9] C. M. Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, 1995 Oxford, UK.

[10] J. Joachims, "Making large-scale SVM learning practical". B. Scholkopf, C. J. C. Burges and A. J. Smola, editors, "Advances in Kernel Methods-Support Vector Learning", 169-184, MIT Press, 1999, Cambridge, MA.

[11] R. D. Reed, and R. J. Marks II, "Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks", Cambridge, MA: The MIT Press, 1999, ISBN 0-262-18190-8.

[12] D. E. Rumelhart, G. E. Hinton, , and R. J. Williams, "Learning internal representations by error propagation", in D. E. Rumelhart, and J. L. McClelland, eds. "Parallel Distributed Processing: Explorations in the Microstructure of Cognition", Volume 1, 318-362, 1986, Cambridge, MA: The MIT Press.

[13] S. Schwarm, I. Bulyko, and M. Ostendorf, "Adaptive Language Modeling with Varied Sources to Cover New Vocabulary Items", IEEE Transactions on Speech and Audio Processing, May 2004, vol. 12, n. 3.

[14] B. Bigi, Y. Huang, and R. Mori, "Vocabulary and Language Model Adaptation using Information Retrieval", 2004, In Proceeding. of ICSLP.

[15] G. Boulianne, et al., "Computer-assisted closed-captioning of live TV broadcast in French", 2006, In Proceeding of Interspeech.

[16] L. Chen, J. Gauvain, L. Lamel, and G. Adda, "Dynamic Language Modeling for Broadcast News", 2004, In Proceeding of ICSLP.

[17] IIT, UST Bannu, N-W.F.P, 2006.