Classification of Breast cancer by comparing Back propagation training algorithms

F.Paulin
Asst. Professor, Department of Computer Science
CMS College of Science & Commerce
Coimbatore, Tamilnadu, India

A.Santhakumaran
Professor of Mathematics
Indian Institute of Crop Processing Technology
Tanjore, Tamilnadu, India

Abstract— Breast cancer diagnosis has been approached by various machine learning techniques for many years. This paper presents a study on classification of Breast cancer using Feed Forward Artificial Neural Networks. Back propagation algorithm is used to train this network. The performance of the network is evaluated using Wisconsin breast cancer data set for various training algorithms. The highest accuracy of 99.28% is achieved when using levenberg marquardt algorithm.

Keywords-Breast cancer; Back propagation algorithm; Quasi-Newton.

I. INTRODUCTION

Breast cancer is cancer of breast tissue. Breast cancer is a malignant tumor that has developed from cells of the breast. Breast cancer has become a major cause of death among women in developed countries [4]. The most effective way to reduce breast cancer deaths is detect it earlier. However earlier treatment requires the ability to detect breast cancer in early stages. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to distinguish benign breast tumors from malignant ones. The automatic diagnosis of breast cancer is an important, real-world medical problem [16]. Thus, finding an accurate and effective diagnosis method is very important. The breast cancer diagnosis problem has attracted many researchers in computational intelligence, data mining, and statistical fields. Artificial neural networks (ANNs) [6] have been recently proposed as a very effective method for pattern recognition, machine learning and data mining. ANNs are massively distributed parallel processing systems made up of highly interconnected neural computing elements that have the ability to learn and thereby acquire knowledge and make it available for use [3]. To enable an automatic design of ANNs and perform a global search, evolutionary ANNs have been widely explored in recent years [1]. Neural networks have been widely used for breast cancer diagnosis [13], [14]. Feed forward neural networks (FFNN) are commonly used for classification. Feed forward neural networks have been trained with standard Back propagation algorithm [11]. They are supervised networks so they require a desired response to be trained. They have been shown to approximate the performance of optimal statistical classifiers in difficult problems.

Wisconsin Breast Cancer Data (WBCD) is analyzed by various researchers on medical diagnosis of breast cancer in neural network literature [7],[8],[10],[12],[16]. In [2] Breast cancer is diagnosed using feed forward neural networks by comparing the hidden neurons. In [5], the performance comparison of the multilayered perceptron networks using various back propagation algorithms for breast cancer diagnosis is discussed. The training algorithms used are gradient descent with momentum and adaptive learning, resilient back propagation, Quasi-Newton and Levenberg-Marquardt. The performances of these four algorithms are compared with the standard steepest descent back propagation algorithm. The MLP network using the Levenberg-Marquardt algorithm displays the best performance. The seventh attribute called Bare Nuclei of WBCD has 16 missing values. In [2], the 16 missing value instances have been left out while using WBCD for Breast Cancer diagnosis. The constructed feed forward neural network has been evaluated for breast cancer detection without replacing missing values [9]. Eliminating some instances will affect the diagnosis accuracy. This paper presents a result of direct classification of data after replacing missing values using median method for the WBCD dataset with various back propagation training algorithms. The training algorithms are compared using accuracy.

This paper is organized as follows. Section II contains methodology. Section III gives an application. Finally, the conclusions are drawn in Section IV.

II. METHODOLOGY

The development of artificial neural networks was initially motivated by research into biological nervous systems which consist of densely connected networks of neurons. In the human nervous system, there are over 100 types of neurons. .

The basic structure of the neural network in this paper is a multi-layered feed forward neural network. The number of input units corresponds to the features of the given examples. The number of the output units is determined by the result. The number of hidden units depends on the problem in hand. The activation function of each unit is a sigmoid function. The artificial network is trained by using the standard back propagation algorithm.

Feed forward networks feed outputs from individual neurons forward to one or more neurons or layers in the network. The output of a node is scaled by the connecting weight and is fed forward as an input to the nodes in the next layer of the network. The input layer plays no computational role but merely serves to pass the input vector to the network. The input layer and the hidden layer are connected by weights and likewise the hidden layer and output layer also have connection weights. The network has the ability to learn through training. The training requires a series of input and associated output vectors. During the training, the network is repeatedly presented with the training data and the weights and thresholds in the network are adjusted from time to time till the desired input output mapping occurs.

Artificial neuron performs the following.

1. Receives signal from other neurons.

2. Multiplies each signal by the corresponding connection strength, that is weight.

3. Sums up the weighted signals and pass them through an activation function.

4. Feeds output to other neurons.

Usually the final layer neurons can have linear activation functions while intermediate layer neurons implement nonlinear functions, since most real world problems are nonlinearly separable, nonlinearity in intermediate layers is essential for modeling complex problems.

A. Data Normalization

One of the most common tools used by designers of automated recognition systems to obtain better results is to utilize data normalization. Ideally a system designer wants the same range of values for each input feature in order to minimize bias within the neural network for one feature over another. Data normalization can also speed up training time by starting the training process for each feature within the same scale.

Input data has been normalized by the min-max normalization as in (1), in the range between 0 and 1:

$$\overline{X} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \qquad (1)$$

Where, $\overline{X}$ is standard value of input,

x is Observed value,

$x_{\max}$, $x_{\min}$ are maximum and minimum actual observed values.

The following pre-classification rule have adopted in this work. In which three fields are included: Clump thickness, Bare Nuclei, and Mitoses as given below.

If (Clump thickness < 7 and Uniformity of cell size<8 and Uniformity of cell shape<3 and Normal Nucleoli<9) then

Benign

Else

Malignant

B. Replacement of missing values using Median Method

1. Find median for the Bare Nuclei (This attribute contains missing values). The median is calculated using the formula,

$$Median = sizeof \frac{(N+1)}{2} \text{ th item.}$$

2. All the missing value of this attribute replaced by this median value.

There are generally four steps in training process:

1. Assemble the training data
2. Create the network object
3. Train the network
4. Simulate the network response to new inputs

There are many variations of the back propagation algorithm, several of which are described here.

C. Batch Training

In batch mode the weights and biases of the network are updated only after the entire training set has been applied to the network. The gradients calculated at each training example are added together to determine the change in the weights and biases.

D. Batch Gradient Descent

This is the batch steepest descent training algorithm. The weights and biases are updated in the direction of the negative gradient of the performance function.

E. Batch Gradient Descent with Momentum

In provides faster convergence, steepest descent with momentum. Momentum allows a network to respond not only to the local gradient, but also to recent trends in the error surface. Acting like a low pass filter, momentum allows the network to ignore small features in the error surface. Without momentum a network can get stuck in a shallow local minimum.

F. Conjugate Gradient Algorithms

The basic back propagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient), the direction in which the performance function is decreasing most rapidly. It turns out that, although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. In the conjugate gradient algorithms a search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions.

G. Quasi-Newton Algorithms

Newton's method is an alternative to the conjugate gradient methods for fast optimization. There is a class of algorithms that is based on Newton's method, but which doesn't require calculation of second derivatives. These are called quasi-Newton (or secant) methods. They update an approximate Hessian matrix at each iteration of the algorithm. The update is computed as a function of the gradient.

H. Levenberg-Marquardt

Like the quasi-Newton methods, the Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix.

I. Resilient Back propagation

Multilayer networks typically use sigmoid transfer functions in the hidden layers. These functions are often called "squashing" functions, because they compress an infinite input range into a finite output range. Sigmoid functions are characterized by the fact that their slopes must approach zero as the input gets large. This causes a problem when you use steepest descent to train a multilayer network with sigmoid functions, because the gradient can have a very small magnitude and, therefore, cause small changes in the weights and biases, even though the weights and biases are far from their optimal values.

The purpose of the resilient back propagation training algorithm is to eliminate these harmful effects of the magnitudes of the partial derivatives. Only the sign of the derivative is used to determine the direction of the weight update; the magnitude of the derivative has no effect on the weight update. The size of the weight change is determined by a separate update value.

The Back-propagation algorithm and supervised training method are used in this project. The aim of training is to adjust the weights until the error measured between the desired output and the actual output is reduced. The training stops when this reaches a sufficiently low value. To analyze the data neural network tool box which is available in MATLAB software is used.

The proposed algorithm used in this research is follows.

1. Load data set and replace missing values by using Median method

2. Normalize the data using min-max normalization

3. Create a network using Back propagation algorithm with an input layer, a hidden layer and an output layer. There are 9 neurons in the input layer, 6 neurons in the hidden layer and 1 neuron in the output layer.

4. Assign random value to the input weight and the bias.

5. Allocate 80% for training and the remaining 20% for testing.

6. Train and test the network using Back propagation algorithm with various training algorithms.

7. Compare the results of the training algorithms using accuracy.

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this file and download the file for "MSW A4 format".

III. An Application

A. Wisconsin Breast Cancer Database

The Wisconsin breast cancer database was originally provided by Dr. William H. Wolberg [15] and used by a number of researchers in pattern recognition and machine learning. The database available in the UCI database repository contains 699 cases. The original dataset contains 11 attributes including both sample id number and class label, which are removed in the actual dataset that are used in this application. The class of each instance is either benign or malignant. The remaining 9 attributes represent 9 cytological characteristics of breast fine-needle aspirates (FNAs), as shown in TABLE I. The cytological characteristics of breast FNAs were valued on a scale of one to ten, with one being the closest to benign and ten the most malignant.

TABLE I. Attribute information

No.	Attribute	Domain
1	Clump thickness	1-10
2	Uniformity of cell size	1-10
3	Uniformity of cell shape	1-10
4	Marginal Adhesion	1-10
5	Single Epithelial cell size	1-10
6	Bare Nuclei	1-10
7	Bland Chromatin	1-10
8	Normal Nucleoli	1-10
9	Mitoses	1-10
10	Class	2 for benign and 4 for malignant

The data set was partitioned into two sets: training and testing set. The testing set was not seen by any neural network during the training phase. It is only used for testing the generalization of neural network ensembles after they are trained. We used the 80% examples for the training set, and the rest 20% examples for the testing set. There are three layers in the back propagation network, including an input layer containing nine units, a hidden layer containing six units and an output layer containing only one unit. The value of the unit in the output layer shows whether the input is a normal cell or not. At the start of training, all connection weights in network are set to random values. All input vectors are normalized so that the minimum and maximum are 0 and 1 respectively. All the computations are implemented using MATLAB V7. The function newff was used to create the feed forward Back propagation network. The architecture used in these applications consisted of tan-sigmoid hidden units and one purelin output unit. The learning rate of 0.7 was used. Number of maximum allowable epochs was 1000.

TABLE II shows the effect of training algorithms on the accuracy of diagnosis with Back Propagation Algorithm. From this table, it is clear that the Feed Forward Neural Network using Back Propagation Algorithm with Levenberg Marquardt training algorithm produced the best results for diagnosis.

TABLE II. PERFORMANCE OF THE TRAINING ALGORITHMS

Name of the Training Algorithm	**% of Accuracy of Diagnosis**
Batch Gradient Descent (BGD)	83.27
Batch Gradient Descent with Momentum (BGDM)	84.39
Quasi Newton (QN)	98.42
Resilient Back Propagation (RBP)	98.63
Conjugate Gradient (CG)	98.99
Levenberg Marquardt (LM)	99.28

Fig. 1 illustrates the chart of relationship between the number of hidden layers and the percentage of correct classification.

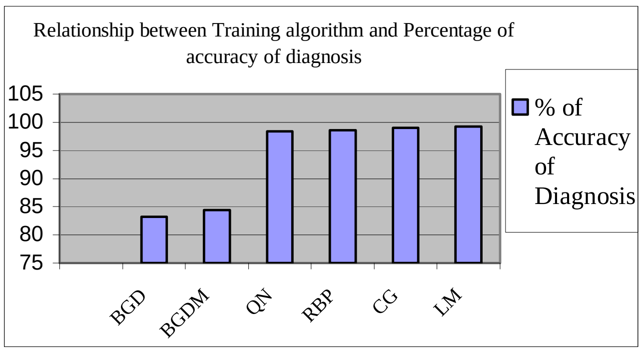

Figure 1. Relationship between the training algorithms and the percentage of accuracy of diagnosis

IV. CONCLUSION

In this research a feed forward neural network is constructed and the Back propagation algorithm is used to train the network. The proposed algorithm is tested on a real life problem, the Wisconsin Breast Cancer Diagnosis problem. In this paper six training algorithms are used, among these six methods, Levenberg Marquardt method gave the good result of 99.28%. Preprocessing using min-max normalization is used in this diagnosis. Further work is needed to increase the accuracy of classification of breast cancer diagnosis.

REFERENCES

[1] Abraham, A. “Meta learning evolutionary artificial neural networks”, Neuro computing, 2004, 56, pp.1-38.

[2] Anupam Shukla, Ritu Tiwari and Prabhdeep Kaur, “Knowledge Based Approach for Diagnosis of Breast Cancer” IEEE International Advance Computing Conference, Patiala, India, March 2009, pp. 6-12

[3] Basheer I.A., and Hajmeer M. “Artificial neural networks: fundamentals, computing, design, and application” J. Microbiol. Methods 43 (2000) 3-31.

[4] Bothorel S., Meunier B.B., and Muller S.A., “Fuzzy logic based approach for semi logical analysis of micro calcification in mammographic images” International Journal of Intelligent System, 12 (1112): 819-848, 1997.

[5] Esugasini S, Mohd Yusoff Mashor, Nor Ashidi Mat Isa and Nor Hayati Othman (2005), Performance Comparison for MLP Networks Using Various Back Propagation Algorithms for Breast Cancer Diagnosis, Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, 3682, pp. 123-130.

[6] Furundzic D., Djordjevic, and Bekic A. J., “Neural Networks approach to early breast cancer detection”, Systems Architecture, 44: 617- 633, 1998.

[7] Jun Zhang MS, Haobo Ma Md MS, “An Implementation of Guildford Cytological Grading System to diagnose Breast Cancer Using Naïve Bayesian Classifier”, MEDINFO 2004, M.Fieschi et al. (Eds),Amsterdam:IOS Press

[8] Kamruzzaman S.M. and Monirul Islam Md, “Extraction of Symbolic Rules from Artificial Neural Networks” Proceedings of world Academy of science, Engineering and Technology, vol. 10, Dec. 2005, ISSN 1307-6884

[9] Paulin F. and Santhakumaran A., “Extracting Rules from Feed Forward Neural Networks for Diagnosing Breast Cancer” CiiT International Journal of Artificial Intelligent Systems and Machine Learning, vol. 1, No. 4, July 2009, pp. 143-146

[10] Punitha A., Sumathi C.P. and Santhanam T., “A Combination of Genetic Algorithm and ART Neural Network for Breast Cancer Diagnosis” Asian Journal of Information Technology 6 (1):112-117, 2007, Medwell Journals, 2007.

[11] Renato De Leone, Rosario Capparuccia and Emanuela Marelli, “A Successive Overrelaxation Backpropagation Algorithm for Neural-Network Training” IEEE Transactions on Neural Networks, vol. 9, No. 3, May 1998, pp. 381-388

[12] Rudy Setiono and Huan Liu, “Neural-Network Feature Selector” IEEE Transactions On Neural Networks, vol. 8, No. 3, May 1997, pg 664-662

[13] Tuba Kiyan And Tulay Yildirim, “Breast Cancer Diagnosis Using Statistical Neural Networks” Istanbul University, Journal Of Electrical And Electronics Engineering, Year 2004, vol. 4, Number 2, pp.1149-1153

[14] Sudhir D. Swarkar, Ashok Ghatol, Amol P. Pande, “Neural Network Aided Breast Cancer Detection and Diagnosis Using Support Vector Machine” Proceedings of the International conference on Neural Networks, Cavtat, Croatia, June 12-14, 2006, pp. 158-163.

[15] UCI Machine Learning Repository. [http://archive.ics.uci.edu/ml/]. Irvine, CA: University of California, Center for Machine Learning and Intelligent Systems.

[16] Wlodzislaw Duch and Rafal Adamczak and Krzysztof Grabczewski, “A New methodology of Extraction, Optimization and Application of Crisp and Fuzzy Logic Rules” IEEE Transactions On Neural Networks, vol. 12, No. 2, March 2001, pp. 227-306