

Back Propagation Neural Network by Comparing Hidden Neurons: Case study on Breast Cancer Diagnosis

F. Paulin

Senior Lecturer, CMS College of
Science and commerce, Coimbatore,
Tamilnadu, India.

Dr.A.Santhakumaran

Associate Professor and Head,
Department of statistics, Salem
Sowdeswari College, Salem-10,
TamilNadu, India.

ABSTRACT

This paper investigates the potential of applying the feed forward neural network architecture for the classification of breast cancer. Back-propagation algorithm is used for training multi-layer artificial neural network. Missing values are replaced with median method before the construction of the network. This paper presents the results of a comparison among ten different hidden neuron initialization methods. The classification results have indicated that the network gave the good diagnostic performance of 99.28%.

Keywords

Artificial Neural Networks, Back propagation, Breast cancer, Median.

1. INTRODUCTION

A major class of problems in medical science involves the diagnosis of disease, based upon various tests performed upon the patient. When several tests are involved, the ultimate diagnosis may be difficult to obtain, even for a medical expert. This has given rise, over the past few decades, to computerize diagnostic tools, intended to aid the physician in making sense out of the confusion of data. Breast cancer is one of the leading cancers for women worldwide. The automatic diagnosis of breast cancer is an important, real-world medical problem [11]. Breast Cancer forms in tissues of the breast, usually the ducts (tubes that carry milk to the nipple) and lobules (glands that make milk). A breast lump is a mass within the breast. Fluid-filled masses are known as cysts. Solid masses may be non cancerous tumors, or they may be breast cancer. Breast cancer occurs when a cell within a breast undergoes changes that cause it to grow and divide uncontrollably, that follow some simple guidelines. If any mass or lump is thought to be suspicious, a diagnostic mammogram is usually performed. If the mammogram suggests the possibility of cancer, a breast biopsy, a tissue sample is removed and checked for cancer. If the mammogram shows a fluid-filled cyst, it can be drained with a needle. Depending on the results, further steps may be taken [16]. Fine Needle Aspiration (FNA) is a percutaneous (“through the skin”) procedure that uses a fine gauge needle (22 or 25 gauge) and a syringe to sample fluid from a breast cyst or remove clusters of cells from a solid mass. With FNA, the cellular material taken from the breast is usually sent to the pathology laboratory for analysis [16]. Early detection and improved therapy planning are crucial for increasing the survival rates of cancer patients. To aid clinicians in the diagnosis of breast cancer, recent research has looked into the development of computer aided diagnostic tools.

Neural networks have been widely used for breast cancer diagnosis [10], [11]. Artificial Neural Network (ANN) has made a significant mark in the domain of health-care applications. ANNs are highly structured information processing units operating in parallel and attempting to mimic the huge computational ability of the human brain and nervous system by attempting to emulate the human brain, neural networks learn from experience, generalize from previous examples, abstract essential characteristics from inputs containing irrelevant data. The brain basically learns from experience, In ANN, learning is typically achieved through progressive adjustment of the weighted interconnections of neurons and other network parameters, guided by learning algorithm. Feed forward neural networks (FFNN) are commonly used for classification. Feed forward neural networks have been trained with standard Back propagation algorithm [7]. 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.

There is much research on medical diagnosis of breast cancer with Wisconsin Breast Cancer Data (WBCD) in neural network literature [3],[4],[6],[9],[12]. In [1] Breast cancer is diagnosed using feed forward neural networks by comparing the hidden neurons, the highest accuracy was 99.5%. In [1], 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 [5]. Eliminating some instances will affect the diagnosis accuracy. The seventh attribute called Bare Nuclei of WBCD has 16 missing values. This paper presents a result of direct classification of data after replacing missing values using median method for the WBCD dataset with a given number of classes.

1.1 Data Set

This breast cancer database is downloaded from the UCI machine learning repository [17], which was collected by Dr. William H. Wolberg from the University of Wisconsin Hospitals, Madison [13]. The dataset is comprised of elements that consist of various scalar observations. The total number of the original samples is 699 with 16 samples contain missing values. The dataset contains two classes referring to benign and malignant samples. There are 458 samples in the dataset that are assigned to benign and the other 241 samples are malignant. 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 our experiments. The remaining 9 attributes represent 9 cytological characteristics of breast fine-needle aspirates (FNAs),

as shown in Table 1. 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 1. Attribute Information

S.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

Number of instances: 699

Missing Attributes : 16

Benign : 458

2. METHODOLOGY

In Feed forward neural networks, the neurons are arranged in layers, with the first layer taking in inputs and the last layer producing outputs. The middle layers have no connection with the external world, and hence are called hidden layers. Each neuron in one layer is connected to every neuron on the next layer. Hence information is constantly fed forward from one layer to the next. There is no connection among neurons in the same layer.

Learning in feed forward networks belongs to the realm of supervised learning, in which pairs of input and output values are fed into the network for many cycles, so that the network ‘learns’ the relationship between the input and output. In Back propagation learning, every time an input training vector of a training sample is presented, the output vector o is compared to the desired value d .

The comparison is done by calculating the squared difference of the two as in equation (1):

$$Error = (d - o)^2 \text{ ---- (1)}$$

The value of Err tells how far away from the desired value for a particular input. The goal of back propagation is to minimize the sum of Err for all the training samples, so that the network behaves in the most desirable way.

$$Minimize Error = \sum (d_i - o_i)^2 \text{ ---- (2)}$$

We can express the $Error (Err)$ in terms of the input vector (i), the weight vector (w), and the threshold function of the neurons. Using a continuous function as the threshold function, the gradient of Err with respect to the w in terms of w and i can be expressed.

Given the fact that decreasing the value of w in the direction of the gradient leads to the most rapid decrease in Err , the weight vectors updated every time a sample is presented using the following formula

$$W_{new} = W_{old} \cdot n \frac{\delta Err}{\delta w} \text{ ---- (3)}$$

Where n is the learning rate in equation (3).

Using this algorithm, the weight vectors are modified so that the value of Err for a particular input sample decreases a little bit every time the sample is presented. When all the samples are presented in turns for many cycles, the sum of Err gradually decreases to a minimum value, which is the goal of back propagation algorithm.

The feed forward back propagation neural network can learn a function of mapping inputs to outputs by being trained with cases of input-output pairs. Back propagation neural network (BPNN) is actually a descending slope method to minimize the total square of the output, calculated by the network [8]. There are three phases in the training process: first is to send the signal pattern forward, second is to calculate the propagated error and the last is to update all weights in the network. In addition BPNN also have the advantages of faster learning in multilayer Neural Network, especially sigmoidal activation function is represented by hyperbolic tangent. The neurons in feed forward networks can be any transfer function of the designer wishes to use. The most commonly used transfer function is the sigmoid function as in equation (4).

$$f(x) = \frac{1}{1+e^{-x}} \text{ ----- (4)}$$

where x is the input to the node.

The network performance and convergence depends on many parameters like initial weights, learning rate and momentum used, number of nodes in the hidden layer during the training process. Figure 1 illustrates the flowchart of the overall processes used in this research.

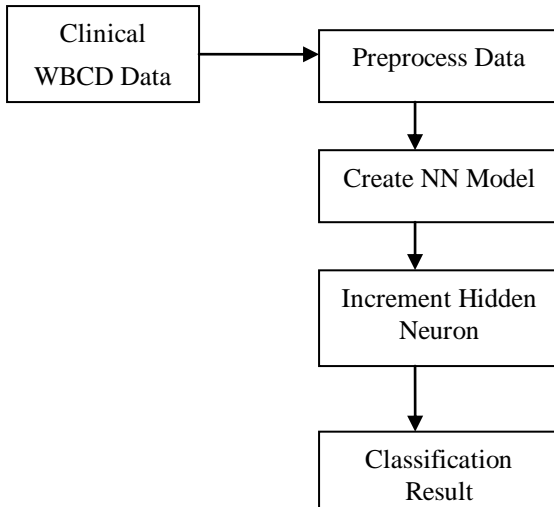


Figure. 1 Overall Processes Flowchart

The proposed algorithm used in this research is follows.

1. Load Data set.
2. Replace missing values by median missing values replacement method.
3. Normalize each variable of the data set, so that the values range from 0 to 1. We call this data set as normalized data set.
4. Create a separate training set and testing set by randomly drawing out 80% of the data for training and 20% for testing.
5. Create an initial ANN architecture consisting of three layers, an input, an output and a hidden layer. The number of nodes in the input layer is the same as the number of inputs of the problem. and the output layer contains 1 node. Increase the number of hidden nodes in hidden layer for every run. Randomly initialize all connection weights within a certain range.
6. Train the network on the training set by using Back propagation algorithm until the error is almost constant for a certain number of training epochs, this is specified by the user.
7. Present the test data to the trained network and evaluate the performance.

2.1 Replacement of missing values using Median Method:

1. Find median for the Bare Nuclei (This attribute contains missing values).
2. All the missing value of this attribute replaced by this median value.

3. AN APPLICATION

Preprocessing the input data set for a knowledge discovery goal using the neural network approach usually consumes the biggest portion of the effort devoted in the entire work. A simple analysis shows that the WBCD data set has missing information in the field of Bare Nuclei for 16 records. In this research, these missing values have been replaced by the calculated value using median replacement method.

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

3.1. Normalize the data

One of the most common tools used by designers of automated recognition systems to obtain better results is to utilize data normalization. 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 formulae as in (2), in the range between 0 and 1:

$$\bar{X} = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (2)$$

Where \bar{X} is the standard value of input,

x is the Observed value,

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

The above mentioned approaches for the proposed algorithm have been implemented and tested with the breast cancer dataset from the University of Wisconsin Hospitals, Madison, collected by Dr. W. H. Wolberg. To analyze the data neural network tool box which is available in MATLAB software is used.

The feed forward neural network (FFNN) consists of an input layer, an output layer and a hidden layer, with 9 features in each input vector and 2 values in each output vector. Initially, a single node is assigned to the hidden layer. After every run the hidden node would be increased by one node, up to 10 nodes in the hidden layer. The training algorithm is a standard back-propagation with a set of initialized parameters. The non-linearity in the hidden layer is the sigmoid function. The Levenberg-Marquardt (trainlm) algorithm was applied to increase the training speed. This trainlm algorithm appears to be the fastest method for training moderate-sized feed forward

neural networks (up to several hundred weights). In this application, 80% of the data were selected randomly and used to train and construct the network. The remaining 20% of the data were then used to test capability of the resulting network. The proposed algorithm was executed 10 times each with different number of neurons in the hidden layer. In Table 2 the percentage of correct classification indicates the percentage of the patterns that were correctly classified by the constructed networks. Table 2. shows the accuracies of different number of hidden neurons.

Table 2. Performance of the Replacement Methods

S. No.	Number of Hidden neurons	Percentage of Correct Classification
1	1	98.63%
2	2	98.70%
3	3	98.92%
4	4	98.92%
5	5	98.85%
6	6	99.28%
7	7	98.20%
8	8	98.92%
9	9	98.92%
10	10	98.78%

The network has given the good diagnostic performance of 99.28% for 6 neurons in the network.

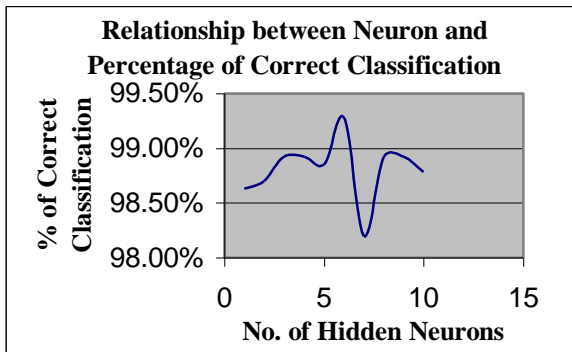


Figure. 2 Relationship between the number of hidden neurons and the percentage of correct Classification

Figure 2 illustrates the graph of relationship between the number of hidden layers and the percentage of correct classification.

4. 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 Median missing value replacement method and preprocessing using min-max normalization are used. Hidden neurons are increased by one neuron after every run up to ten neurons in the network. The network has given the good diagnostic performance of 99.28% for 6 neurons. Further work is needed to increase the accuracy of classification of breast cancer diagnosis.

5. REFERENCES

- [1] Anupam Shukla, Ritu Tiwari and Prabhdeep Kaur, Knowledge Based Approach for Diagnosis of Breast Cancer” IEEE International Advance Computing Conference, Patiala, India, (Mar. 2009), 6-12.
- [2] Bennett K. and Mangasarian O. L., “Neural Network Training via Linear Programming,” Advances in Optimization and Parallel Computing. Elsevier Science Publishers, 1992.
- [3] 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
- [4] 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
- [5] 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, (Jul. 2009), 143-146
- [6] 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, Medwell Journals, 2007.
- [7] 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), 381-388
- [8] Ruck D.W., Rogers S.K., M.Kabrisky., Meibeck P. S., and M. E. Oxley., ”Comparatives Analysis of Backpropagation & the extended Kalman Filter for Training Multilayer perceptrons”, IEEE Transactions on Pattern Analysis and Machine Intelligence, June 1992, Vol 14, No 6, 686-691
- [9] Rudy Setiono and Huan Liu, “Neural-Network Feature Selector” IEEE Transactions On Neural Networks, vol. 8, No. 3, (May 1997), 664-662
- [10] 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, 158-163.
- [11] Tuba Kiyani And Tulay Yildirim, "Breast Cancer Diagnosis Using Statistical Neural Networks" Istanbul University, Journal Of Electrical And Electronics Engineering, Year 2004, vol. 4, Number 2, 1149-1153
- [12] 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, (Mar. 2001), 227-306
- [13] Wolberg W. H. and Mangasarian O.L, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology," in Proceedings of the National Academy of Sciences, vol. 87, 9193-9196, U.S.A., *(Dec. 1990).
- [14] American Cancer Society, Cancer Facts and Figures 2007, 2007th ed. American Cancer Society, 2007.
- [15] <http://www.healthon.com/topics/B/2/Breast-Masses-and-Cancer/Breast-Lumps-and-Cancer...>
- [16] <http://www.imaginis.com/breasthealth/biopsy/fine.asp>
- [17] UCI Machine Learning Repository. [<http://archive.ics.uci.edu/ml/>]. Irvine, CA: University of California, Center for Machine Learning and Intelligent Systems.
- [18] U. S. Cancer Statistics Working Group, "United states cancer statistics 2003 incidence and mortality (preliminary data)," National Vital Statistics, vol. 53, no. 5, 2004.