

Breast Cancer Detection using ART2 Model of Neural Networks

Sonia Narang¹, Harsh K Verma², Uday Sachdev³

Department of Computer Science & Engineering
Dr. B.R. Ambedkar National Institute of Technology, Jalandhar (India)

ABSTRACT

Breast cancer has become a major cause of death among women. To reduce breast cancer deaths the most effective way is to detect it earlier. Early diagnosis requires correct and reliable diagnosis procedure that will allow physicians to distinguish benign breast tumors from malignant ones. So, to find an accurate and effective diagnosis method is very important.

This paper presents a system which detects the cancer stage using Adaptive Resonance theory (ART2) Neural Network. A vigilance parameter (vp) in ART2 defines the stopping criterion and hence helps in manipulating the accuracy of the trained network. We see that at vp=0.2 the network has Recall (i.e. true negative rate) is 75% and average Accuracy=82.64% and Precision is 79%.

Keywords – ARNN, ART1, ART2, Neural Networks, Breast Cancer Detection.

1. INTRODUCTION

Adaptive resonance theory (ART) [1] models are a neural network [4][5] that does clustering, and in addition they can offer the number of clusters to vary with the size of the problem. The major deviation between ART2 and former clustering technique is that ART2 empowers the user to command the degree of similarity between members of the same cluster by means of a user defined constant called the vigilance parameter. It applies a MAXNET. A “maxnet” is a recurrent one-layer network that performs a contest to determine which node has the largest initial input value. Apart from neural plausibility and the desire to do every task by applying neural networks, it can be reasoned that the maxnet allows for greater parallelism in performance than the algebraic solution.

The presented system detects whether a patient's cancer stage is malignant or benign [1]. Neural networks are useful to execute this task [2][5][6]. A lot of cancer detection methods depend on historical data. Experiments indicate that a cancer case currently under observation is malignant if it has the same or similar attributes, as a malignant case that has been detected as malignant earlier in the identical conditions. Consequently, historical information assists us to detect whether a case is malignant or benign. Various modeling methods have been proposed for and applied to breast cancer detection [5][6][7][8]. In neural networks historical data that is obtained by regression analysis or by any other means are given to network and the network gets trained consequently. This data is called training data. When we input operational data, the neural network responds accordingly to the trained data. Hence we used Adaptive

Resonance Neural Network model (ART2) which clusters already known modules that are faulty and fault-free. Because it works under unsupervised learning so there is no difference between training data and operational data.

The data which we have used is from UCI machine learning data repository named as ‘Wisconsin Breast Cancer Database’ [3] which have 9 attributes of 699 patients under survey. On the basis of these 9 attributes cases are termed as benign or malignant. The ART2 clusters cases which are benign in one cluster and malignant in another cluster.

This paper presents a new rule ordering and evaluation algorithm which orders extracted rules based on three performance measures to enable them that they can be used by any generic interface engine. Additionally, it provides an integration algorithm to scrutinize the network's output as well as that derived rule base subsystem and gives a final decision, along with an associated confidence measure.

The Wisconsin breast cancer database which is used to train three different feed forward artificial neural networks then three different rule extraction techniques, in addition the rule ordering and integration technique are used to extract the rules from these networks. Experimental results imply that the overall system offers superior generalization performance along with data interpretation even where no prior domain knowledge is available.

2. ART MODEL

Adaptive resonance theory (ART) models are neural networks that perform clustering, and permit the number of clusters to vary the size of the problem. The major deviation between ART and former clustering methods is that ART permits the user to manipulate the degree of similarity between members of same cluster is defined as a user-defined constant called the vigilance parameter. ART networks can be used for many pattern recognition tasks, such as automatic target recognition and seismic signal processing.

2.1 Basic Architecture

The basic architecture of adaptive resonance neural network involves three groups of neurons.

1. Input processing field- F1 layer.
2. Cluster units – F2 layer.
3. Reset mechanism- That controls the degree of similarity of patterns placed on same cluster.

2.2 Input Processing (F1 Layer):-

It is divided into two portions. They are given as follows:-

- (1) Input portion
- (2) Interface portion

The Input part just represents the input vector given, but a lots of processing occurs in this portion in ART2 network.

The Interface portion aggregates the signal from the input portion with the use of F2 layer. That signal is used in comparing the similarity of the input signal to the weight vector for the cluster unit which is chosen for learning. F1 layer is connected to F2 layer through bottom up weights b_{ij} and F2 layer is connected to F1 layer through top down weights t_{ij} .

2.3 Cluster Units (F2 Layer):-

This is competitive layer. The cluster unit that has the largest net input is selected to learn the input pattern. The activations of all other F2 units are set to zero. The interface units now aggregate information from the input and cluster units.

Reset mechanism: - The information from the input units is Aggregated in the interface units. Depending on the similarity between the top-down weight and the input vector, the cluster unit may or may not be allowed to learn the pattern. It is done at the reset unit, based on the signals it receives from the input and interface portions of F1 layer. If cluster unit is not permitted to learn, it is inhibited and a new cluster unit is selected as the candidate.

2.4 Reset Mechanism States:-

There are three states .

- (1). Active:- “ON”. The unit in F2 is on. Its activation is given by d , where $d=0$ for ART2.
- (2) Inactive: - “OFF” the unit in F2 is off. Its activation $=0$, but it is available to participate in the next competition.
- (3) Inhibit:-“OFF”.The unit in F2 is off. Its activation $=0$, and is prevented from participation in further competition during the presentation of the current input vector.

2.5 ART2 Network

ART2 accepts continuous valued vectors. ART2 has highly complex F1 units. The F1 units if ART2 posse's combination of normalization and noise suppression, along with the comparison of weights needed for reset mechanism. ART2 has two types of continuous valued inputs. One is called noisy binary signal and the other truly continuous. The first one can operate with the fast learning type data. The second type of data is more suitable with the slow learning mode.

2.6 Architecture of ART2:-

From the architecture it is seen that, the F1 layer has six types of units (W, X, U, P and Q). Between all the units say, W and X, P and Q, V and U there exists supplemental

unit, which receives signals from respective units, calculates its norms and sends to the other units.

The architecture of ART2 network is shown below;-

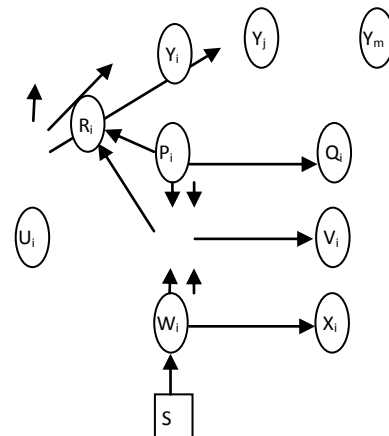



Fig1. ART 2 architecture

The receiving unit receives both inhibitory and excitatory signals from the sending units through supplemental units. The X and Q units are connected to V units. The transformation occurring in the signal is indicated in connection. The P unit path is connected to the cluster units by bottom up and top down weights. The arrow  indicates normalization. The units perform the operation of F1 layer and P units perform the operation of F2 layer interface portion.

2.7 Algorithm:-

The ART2 algorithm involves few differential equations for training. Here w_{ij} , x_i , u_i , v_i , p_i , q_i and y_i are called Short Term Memories (STM) and b_{ij} and t_{ji} are called Long Term Memories (LTM).

2.8 Description

A learning trail consists of one presentation of one input pattern. When the learning trail starts, all activations are set to zero.

The computation cycle starts with the computation of activation of unit U_i . Then the signal is sent from U_i to units W_i and P_i . The activations of W_i and P_i are then computed. W_i sums the signals from U_i and S_i . P_i sums the signals from U_i and top-down signal. X_i and Q_i activations are the normalized forms of W_i and P_i respectively. Before signal is sent to V_i , activation is calculated on each unit. Finally V_i sums the signals it receives from X_i and Q_i . This entirely forms one cycle updation of F1 layer.

When the activations of F1 layer reaches equilibrium, P units send their signals to F2 layer, and winner unit is selected based on competition. The reset mechanism checks for reset whenever it receives a signal from P, since the further computations are based on the value of that signal. This signal is going to be the most recent signal the unit R_i had received from U_i . After the check for reset is finished, the cluster unit may be rejected or accepted.

Based on this, the learning process starts. ART2 perform slow learning and fast learning. In slow learning, only one iteration of the weight update equations occurs on each

trail and in fast learning, the weight updates continue until the weights reach equilibrium on each trail. In fast learning, the placement of cluster stabilizes, but the weights will change for each pattern presented.

2.9 Calculations of F1 layer

These calculations are necessary whenever “update F1 activations” occur in the algorithm. The normalization and noise suppression is based on these calculations. ‘J’ indicates the winning unit of the F2 layer based on the competition. If no winning unit is chosen then d=0 for all units. The calculations of w_i and p_i , x_i and q_i can be performed in parallel.

The calculations involved are:

$$u_i = \frac{v_i}{e + v}$$

$$w_i = s_i + au_i, p_i = u_i + dt_{jt}$$

$$x_i = \frac{w_i}{e + w}, q_i = p_i \frac{p_i}{e + p}$$

$$v_i = f(x_i) + bf(q_i)$$

The activation function is,

$$f(x) = \begin{cases} x & \text{if } x \geq \theta \\ 0 & \text{if } x < \theta \end{cases}$$

2.10 Training Algorithm

The training algorithm is as follows:

Step1 : Initialize parameters a,b,θ,c,d,e,α,p.

Step2: Perform steps 3-13 up to specified number of epochs of training.

Step3: For each input vector ‘s’, do step 4-12

Step4: Update F1 unit activations

$$u_i = 0, x_i = \frac{s_i}{e + s}$$

$$w_i = s_i, q_i = 0$$

$$p_i = 0, v_i = f(x_i)$$

Update F1 unit activation again.

$$x_1 = \frac{v_i}{e + \dots}, w_i = s_i + au_i$$

$$p_1 = u_1, x_1 = \frac{w_i}{e + w}, q_1 = \frac{p_i}{e + p}$$

$$v = f(x_i) + bf(q_i)$$

Step5 : Compute signals to F2 units

$$y_i = \sum b_{ij} p_i$$

Step6: While reset is true, perform Steps 7-8.

Step7: For F2 unit choose Y_j with largest signal

Step8: Check for reset

$$u_i = \frac{v_i}{e + v}, p_i = u_i + dt_{jt}$$

$$r_i = \frac{u_i + cp_i}{e + u + c + p}$$

$$\text{if } |r| < p - e, \text{ then}$$

$$y_i = -1(\text{inhibit } j)$$

Since reset is true, go to step6

$$\text{if } |r| > p - e, \text{ then}$$

$$w_i = s_i + au_i$$

$$v_i = f(x_i) + bf(q_i)$$

Reset is false, so go to step9.

Step9: Perform steps 10-12 up to specified number of learning iterations.

Step10: Update weights for winning unit j.

$$t_{ij(\text{new})} = \alpha du_i + \{1 + \alpha d(d - 1)\} t_{ij(\text{old})}$$

$$b_{ij(\text{new})} = \alpha du_i + \{1 + \alpha d(d - 1)\} b_{ij(\text{old})}$$

Step11: Update F1 activations.

(Formulae mentioned in calculations of F1 layer are used)

Step12: Test stopping condition for weight updates.

Step13: Test stopping condition for number of epochs.

In slow learning, number of learning iterations is 1. In fast learning, for the first pattern learned by cluster ue, will be parallel to t throughout training cycle and the equilibrium weights are,

The stopping condition may be number of epochs or the weight changes below a specified tolerance.

$$t_{ij} = \frac{1}{1 - d} u_i, b_{ij} = \frac{1}{1 - d} u_i$$

The parameters used in algorithm can have following values.

Table1.Parameters used in ART2

Parameters	Description
n	Number of input units
m	Number of output units
a	Fixed weight in F1 layer (10)
b	Fixed weight in F1 layer (10)
c	Fixed weight used in testing for reset (0.1)
d	Activation of winning F2 unit (0.9)
e	Parameter to prevent division by zero when norm of vector is zero (negligible value)
θ	Noise suppression parameter (1/sqrt(n))
α	Learning rate (small value)
ρ	Vigilance parameter (small value)
$t_{ij}(0)$	Initial top-down weight (0)
$b_{ij}(0)$	Initial bottom up weight

3.PERFORMANCE AND FUTURE SCOPE

In this software, cancer stage prediction has been initiated to produce information for early detection of cancer stage and to obtain a reliable diagnostic process. A large number of empirical studies on various aspects of cancer detection are available and several of these incorporate data from industrial projects. Up to now, few surveys actually provide insight on how cancer detection can be applied. In the field of neural networks only classification is applied to cancer detection dataset. Among these are studies from F.Paulin who had used feed forward neural networks to classify the dataset as benign and malignant.

Now we have to predict the performance of our model of 'Breast Cancer Detection Using Adaptive Resonance Neural Networks (ARNN). We are using clustering technique to cluster the dataset into two groups (benign or malignant). It is a binary prediction model, so the performance measures of binary models can be used to predict its performance.

On the basis of results produced by the software made a performance analysis has been conducted by giving an analogy of software defect prediction. It is given as follows:-

Typically the performance of a binary prediction model is summarized by the so called confusion matrix (in case of software defect prediction), which consists of the following four counts:-

1. Number of defective modules predicted as defective (true positive i.e. tp)
2. Number of non-defective modules predicted as defective (false positive i.e. fp)
3. Number of non-defective modules predicted as non-defective (true negative i.e. tn)
4. Number of defective modules predicted as non-defective (false negative i.e. fn)

In case of our modeled software these measures can be defined as follows:-

1. tp = Number of malignant cases predicted as malignant.
2. fp = Number of benign cases predicted as malignant.
3. tn = Number of benign cases predicted as benign .
4. fn = Number of malignant cases predicted as benign.

One may desire tp and tn to have larger values and fp , fn to be smaller but we can measure it using certain parameters defined as follows:-

3.1 Precision:-

The precision measures the chance of correctly predicting malignant cases among the cases classified as malignant. Either a smaller number of correctly predicted malignant cases or a large number of erroneously tagged benign cases would result in a low precision.

3.2 Recall (true positive rate):-

It is often known as true positive rate.

$$recall = \frac{tp}{tp + fn}$$

3.3. Accuracy: -

The accuracy measures the chances of correctly predicting the malignant tendency of individual cases. It ignores the data distribution and cost information. Therefore, it can be misleading criterion if malignant cases are likely to represent a minority of the cases in the dataset.

$$accuracy = \frac{tp + tn}{tp + fn + tn + fp}$$

3.4. True Negative Rate:-

$$True\ negative\ rate = \frac{tn}{tn + fp}$$

3.5 Performance analysis and future scope

In our project we have total number of cases 600 from which 375 are benign cases and 225 are malignant cases. On different values of vigilance parameter (vp) we calculated these parameters and compared them as follows:-

Table 2:- performance of ART2

Vp	tp	fp	tn	fn	precision	Recall/ tp rate	accuracy	tn rate
0.3	165	57	318	60	74.32%	73.33%	80.5%	84.4%
0.2	170	45	330	55	79.06%	75.55%	82.64%	88%
0.4	160	60	315	65	72.72%	71.11%	79.16%	84%
0.5	165	55	320	60	75%	73.33%	80.83%	85.33%

We can see that the maximum Recall of the ARNN model for cancer detection is 75.55% with vigilance parameter 0.2 which is good.

The ARNN model with vigilance parameter 0.22 ensures that the chances of malignant cases will be detected as malignant are good. When we test software for detecting malignant cases we want that all the malignant cases should be detected as malignant but if recall is low then it may happen that a malignant case fall in benign case cluster which will later can produce incorrect results while implementing the software. But in case of ARNN model with vp 0.22 chances of this situation are low. But due to low accuracy than backpropagation neural network (given in table.--) it may happen that some of the benign cases fall in malignant cases cluster but it create no problem when we implement the software as during testing this problem will be solved. But if we want that accuracy should be better than we can just change the vp to 0.2 in which the accuracy and true negative rate is best.

Table 3:- comparison with other models

Model Name	Accuracy
Integrated neural network and fuzzy logic systems	73%
Digital Mammogram Breast Cancer Diagnosis	73.7%
Fuzzy system	80.6%
Adaptive neural network (ART2) (vary according to vigilance parameter)	82.64%

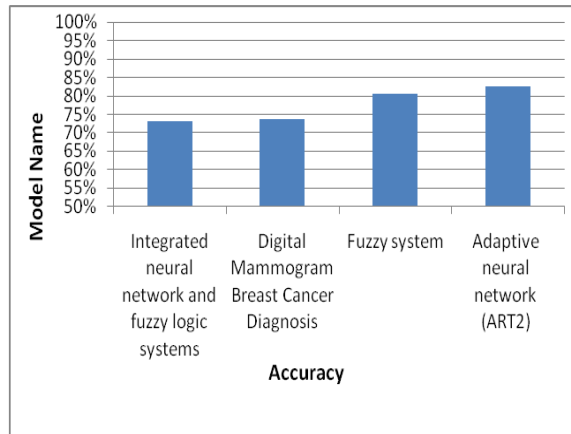


Fig 2:- Comparison with other models

The above information is taken from paper “Breast Cancer Diagnosis Systems: A Review”[11] which compares the performance of various breast cancer diagnosis models.

We can see that the accuracy of ARNN is better as compared to other models but we can enhance the performance by varying the vigilance parameter. This study needs to be extended for the validation using more Datasets. We can also apply the ARNN neural network to other classification problems (for example we have applied ARNN to CRAB CLASSIFICATION which classifies males and female crabs depending on their attributes given in that dataset).

4. CONCLUSION

This dissertation presents a model, cancer detection using adaptive neural network for detecting the cancer stage as benign or malignant. To clusterify the medical data set a neural network approach is adopted. On the Wisconsin Breast Cancer dataset problem, the ART produced good results with better accuracy. We had applied ART2 model of adaptive resonance theory to cluster the dataset as it can be applied on continuous dataset. Clustering of the ARNN expressed reality interpretable knowledge about the dataset, and may be usable by the practitioners in a variety of ways. The result of this model is better the other models in case of accuracy.

The ARNN can be tested on other datasets and variations in the neural networks itself need to be explored. In a developed software system, it would be desirable to improve the accuracy, precision and recall.

5. REFERENCES

- [1]. A.Punitha,C.P.Sumathi and T.Santhanam “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.
- [2]. Xin Yao and Yong Liu,”Neural Networks for Breast Cancer Diagnosis,” International Conference on Neural Networks 2006.
- [3]. Ismail Taha and Joydeep Ghosh “Wisconsin breast cancer database using a hybrid symbolic-connectionist system,” university of Texas Austin 1996.
- [4]. S. N. Sivanandam, Sumathi & Deepa “Introduction to Neural Networks Using Matlab 6.0”. The Mc-Graw hill companies,2006.
- [5]. Kishan Mehrotra, Chilukuri K. Mohan, Sanjay Ranka “Elements of Artificial Neural Networks”. 1997 Massachusetts institute of technology.
- [6]. F.Paulin and A.Santhakumaran, “Classification of Breast cancer by comparing Backpropagation training algorithms,” International Journal on Computer Science and Engineering (IJCSE) 2010.
- [7]. 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, pp.1149-1153
- [8]. Furundzic D., Djordjevic, and Bekic A. J., “Neural Networks approach to early breast cancer detection”, Systems Architecture, 44: 617- 633, 1998.
- [9]. 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. 6-1.
- [10].Abbass HA. An evolutionary artificial neural networks approach for breast cancer diagnosis. Artificial Intelligence in Medicine 2002; 25, 265-281.
- [11].Esugasini Subramaniam, Tan Kuan Liung, Mohd. Yusoff Mashor, Nor Ashidi Mat Isa “Breast Cancer Diagnosis Systems: A Review,”Control and Electronic Intelligent System (CELIS) Research Group,2002