Polynomial Prediction of Neurons in Neural Network Classifier for Breast Cancer Diagnosis

Peter Mc Leod
School of Engineering and Technology
Central Queensland University
160 Ann Street, Brisbane, Queensland, Australia 4000
mcleod.ptr@gmail.com

Brijesh Verma
School of Engineering and Engineering Technology
Central Queensland University
160 Ann Street, Brisbane, Queensland, Australia 4000
b.verma@cqu.edu.au

***Abstract*—Post hoc evaluation mechanisms are utilized for determining the configuration of classifiers. Heuristic approaches mean that sub-optimal configurations could be used; resulting in lost training time, sub-optimal performance and can result in inappropriate results especially for large complex datasets. This paper proposes a new technique to determine the number of neurons in feed forward neural network on two large-scale breast cancer datasets. Classification accuracy of 86% and 89.17% was achieved and the technique predicted the upper and lower bounds for neurons in the feed forward neural networks.**

Keywords—breast cancer, digital mammograms, feed forward neural network

I. INTRODUCTION

Every year research in pattern recognition progresses with the development of novel techniques or the refinement of existing ones. Arguably the most common techniques utilise a neural network that incorporates a number of hidden neurons within one or more hidden layers for classification problems [1]. Although various heuristics exist for determining the number of neurons these rules provide no assurance that an optimal configuration is used resulting in potentially sub-optimal performance. These rules are evaluated post hoc which can mean that an ad hoc approach is taken to training and the cost of training is increased. Ideally we need a mechanism that gives us the ability to define a range or bounds for the selection of neurons to:

- Provide confidence in our selection;
- Reduce unnecessary training time;
- Provide higher or more accurate results;
- Facilitate the development of new novel networks; and
- Allow for the easier addition (retraining) of new knowledge to a classifier [2].

In some instances researchers have determined that finding the optimal configuration is too hard and implement more complex or computationally expensive techniques than required in order to obtain better accuracy. While this is certainly justifiable where the decision boundary is too complex in feature space for other techniques to easily model it is certainly not ideal.

The requirement for determining the most optimal configuration is now more imperative than ever due to the increasing size of datasets and the volume of data to be analysed.

This research evalutes a mechanism for determining the best number of neurons in a feed forward neural network to maximise accuracy. This is performed within the context of a real world problem of breast cancer diagnosis. Although the overriding purpose is to produce highly accurate classifiers,this is done by predicting the number of neurons for a particular network topology (e.g. momentum, iterations) for the feed forward neural network classifiers.

This paper is broken into several sections. Section II covers the background of the research; section III details the research methodology while section IV provides details of the experimental results. Section V discusses the results while conclusions and details about future research are discussed in section VI.

II. BACKGROUND

The purpose behind medical classification is to determine if a patient has the disease or is disease free. Quick and accurate diagnosis provides benefits in the medical sphere by:

- Reducing the load on the hospital system;
- Reduction in psychological stress on patients and families;
- Early diagnosis which can increase the treatment options; and
- Reducing the mortality and morbidity rates.

The use of computer aided diagnostic systems has increased in hospitals in order to effectively realise these benefits. Breast cancer represents a medical clasification

978-1-4673-7679-2/15/$31.00 ©2015 IEEE

paradigm that can be life threatening. In the United States alone a predicted 235,030 new cases of breast cancer are expected in 2014 [3]. Survival rates have improved with the introduction of screening programs and improved treatment regimes. However, a positive prognosis relies on early and accurate detection [4].

According to existing research, the use of CAD systems for screening mammography is now common in countries like the United States having increased from 2% in 2001 to 94% in 2014 [5]. The primary benefits being cited are increases in diagnostic accuracy [6] and reduced diagnosis times [7]. Accordingly ensuring the number of neurons are accurately selected provides major benefits in terms of accuracy and training resources.

A number of researchers have examined the number of hidden neurons in relation to different problem domains. These investigations originally focused on network learning however limited investigation occurred in relation to generalisability and overtraining [8-10]. Arai [11] used a two-parallel hyperplane method to indicate that $2^n/3$ neurons are sufficient for a classification task. Yuan, Xiong and Huai [12] proposed a technique for estimating the number of neurons using decision trees based on information gains when predicting tea quality. The approach required multiple training runs and did not appear to be applied outside their initial tea quality domain. Zhang, Ma and Yang [13] calculated the bounds on three layer neural networks by using the hamming space in a technique called Set Covering Algorithm (SCA). They concluded that $3L/2$ hidden neurons are necessary where L is the number of unit spheres that were contained within the unit sphere covering an n-dimensional hamming space.

Examination of these techniques do not indicate that an evaluation of the learning environment is being incorporated explicitly into the calculation of the number of neurons required. In any classification task, the number of neurons required to perform accurately will be dependent on a number of factors. These will include:

a) The nuances of the classifier (e.g. Learning rate, transfer function, suitability to the classification task e.g. Binary classifier);

b) Problem Domain / Complexity (e.g. Simple / complex problem, discrete outputs, continuous outputs);

c) Training and testing data (number of samples, outliers [17], features for classification);

d) Other factors (e.g. Use of cross validation, restrictions on processing time).

Considering a) to c) above it is very unlikely that a general rule of thumb can be utilised to determine the number of neurons for a classification or regression tasks. The nuances of big datasets that are becoming available in the field require accurate and fast analysis to predict epidemics; pandemics; intervention strategies and discover new meaning within datasets. Thus being able to rapidly determine the optimal configuration is essential to realsing these benefits.

The current levels of accuracy achieved by data mining and intelligent techniques are comparable or better than that of Radiologists. Ubeyli [14] used a support vector machine, neural network, recurrent neural network, probabilistic neural network and multi-layer perceptron to achieve 99.54% classification on breast masses. While Bashir, Qamar and Khan [15] used a technique called heterogeneous classifier which was a weighted based ensemble constituted by five classifiers (Naïve Bayes, Decision Tree using Gini index, Decision Tree using information gain, SVM and a memory based learner) to achieve 97.42% accuracy on the Wisconsin Breast Dataset.

III. Proposed Technique

The ability to be able to predict apriori the performance of a classifier based on the number of neurons is an ideal that would aid in the development of efficient classifiers. Mathematically in it's simplest form this could be written as:

$$A = f(n) \tag{1}$$

Where Accuracy (A) is determined by a function (f) that takes as a parameter being the number of Neurons (n) in a single hidden layer. However the likelihood of achieving this apriori is unlikely due to the factors previously identified. This research proposes that the capacity to predict accuracy within certain bounds exist but is specific to the dataset, the classifier and a number of other factors. This means that the process is not a priori but somewhere in situ.

This research proposes that a number of experiments can be performed at intervals to generate a polynomial regression equation such as that shown below:

$$a=b_n x^n + b_{n-1}x^{n-1} + \ldots + b_2 x^2 + b_1 x + b_0 \tag{2}$$

which can be used to provide a general equation that predicts the approximate accuracy of a neural classifier. We would anticipate that such a regression equation from a number of experiments would tend to produce a parabola of a general form:

$$f(x) = ax^2 + bx + c \tag{3}$$

Thus it would be possible to determine the vertex of such an equation. The verex could be derived by:

$$-b/2a \tag{4}$$

if term b is positive then the parabola would have a convex shape and only one apex. If term b is negative then we would predict a concave shape and potentially we would have two apexes.

Assuming a roughly standard deviation it would be anticipated that a minimum and maximum bound in terms of accuracy would be predicted for feed forward neural network training surrounding the calculated vertex.

Figure I provides an overview of the technique employed in this research.

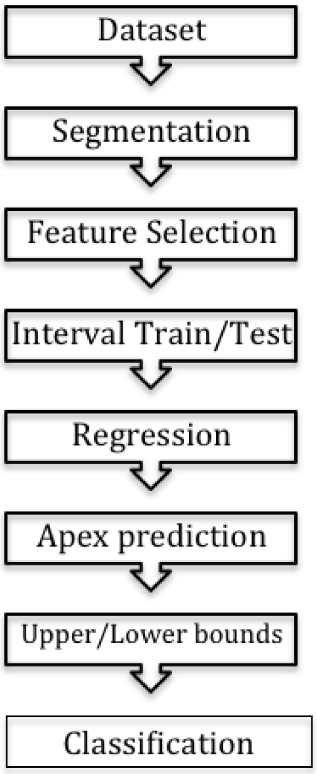

Fig. 1. Proposed technique to determine upper / lower bounds and classify breast cancer anomalies.

A. Dataset (Mammograms)

This research utilises two breast cancer datasets that have been obtained from public benchmark databseses. Table I provides details on the two datasets utilised for this research. The first dataset represents two hundred mass anomalies taken from the Digital Database of Screening Mamography (DDSM) [16]. The DDSM is comprised of 2600+ mammographic cases together with patient information. The second represents 683 fine needle aspirate anomalies taken from the breast cancer dataset from the UCI machine dataset repository [17].

TABLE I. DATASETS UTILISED FOR EXPERIMENTS

Description	Dataset Size	Benign	Malignant
DDSM	200	100	100
UCI	683	444	239

B. Segmentation - ROI (Region of Interest)

Typically the next step is where the mammogram is examined and anomalies are identified. This is then called the Region of Interest (ROI). This process is useful for subsequent phases since the mammogram is divided into ROI which reduces the usage of system resources by discarding inappropriate regions. ROI represent regions of suspicion; however these regions contain anatomical anomalies as well as malignant and begign structures. In this research the area extraction was based on a chain code provided with the DDSM. This process is not required for the UCI dataset.

C. Features

Features in the medical sphere represent observations that map to a diagnosis. A radiologist or clinician would examine the ROIs and make notes according to the ACR BI-RADS® system which is designed to ensure a standardised assessment. The assessment would form part of the patient / clinical notes. Feature selection is one of the most imporant considerations in terms of classification capabilities [18] when mining big data. In this research there are six features used in the DDSM dataset. These are:

- Mass Density;
- Mass Shape;
- Mass Margin;
- Abnormality Assessment Rank;
- Patient Age; and
- Subtlety Value.

There are ten features used in the UCI dataset. These are:

- Radius (mean of distance from centre to points on the perimeter);
- Texture (standard deviation of grey scale values);
- Perimeter;
- Area;
- Smoothness (variation in radius lengths);
- Compactness (perimeter squared / area – 1.0);
- Concavity (severity of concave portions of the contour); and
- Concave Points (Number of concave portions of the countour);
- Symmetry; and
- Fractal Dimension (“coastline approximation” – 1).

D. Interval Train / Test

In order to evaluate if a mechanism could be created to produce a general equation that could predict clasifier accuracy while reducing the cost of training experiments were performed to generate a number of classifiers. The approach involved incrementally growing of a pool of feed forward neural networks occurred containing from 1 to 150 neurons in a single hidden layer at 5 neuron intervals (1,5,10..150) to reduce the search effort. A MATLAB™ seeding function was used to ensure consistency with network initialisation.

The experiments were performed using ten fold cross validation on the same datasets detailed in Table I.

During this time the polynomial regression was being utilised to monitor the performance of the system. Performance was decreasing after approximately 80 neurons were being utilised in the hidden layer but experiments occurred after this point to ensure that the system performance did not increase subsequently.

E. Polynomial Regression

The initial experimental results were analyzed using the polynomial regression described in equation (2). Although an evaluation was performed out to 150 neurons for experimenting since the performance started to decline after 80 neurons. This then allowed for a vertex (apex) to be calculated based on the experiments performed. This vertex represents based on the experimental data the point where the highest accuracy would be predicted to occur.

F. Upper / Lower Bounds

Once an apex was determined it was anticipated that the upper and lower bounds for training a feed forward neural network would be located close to the predicted vertex. This would allow for a degree of confidence when training feed forward neural networks in order to select a higher performing classifier without having to perform a full search across a larger range. Once the apex is determined the lower and upper bounds could be set to determine the actual search space for the best performing classifier.

In this research since experiments were capped at 150 neurons as upper and lower bound of 15 above and below the vertex was chosen as our hypothetical upper and lower bound for the initial evaluation ($1/10^{th}$ of the range utilized).

This upper and lower bound would be analyzed to determine if such bounds were realistic as theoretically more advanced techniques could be utilized based on a normal distribution by calculating the standard distribution.

G. Classification

Following on from the prediction of the upper and lower bounds experiments were performed across the entire range to determine if the highest accuracy fell within the predicted classifier range.

IV. Experimental Results

Details about the configurational paramaters utilised for the feed forward neural networks utilised for this research are recorded in Table II below. Experiments were performed for individual feed forward Multi-Layer Perceptron (MLP) networks (Table III). This allowed the prediction through polynomial regression of a vertex and subsequent experiments were performed to determine across the range to validate if the highest performing classifiers fell within the range.

TABLE II. Neural Network Configuration Details

Paramater	Value
Iterations	3000
Hidden Neurons	1 to 150
Learning Rate	0.7
Momentum	0.5
RMS Goal	0.00001
Transfer Function	Tansig
Network Initialisation	Pseudo-random initilisation

TABLE III. MLP Performance on UCI & DDSM Datasets

Neurons	Dataset	Accuracy [%]
1	UCI	87.70
15	UCI	87.55
30	UCI	88.29
45	UCI	88.58
60	UCI	88.14
75	UCI	87.55
1	DDSM	78.50
15	DDSM	79.00
30	DDSM	85.50
45	DDSM	84.00
60	DDSM	86.00
75	DDSM	84.00

V. Discussion

A polynomial regression was performed to predict the accuracy obtained for different configurations. The regression equation for the two datasets follows below in Table IV.

TABLE IV. Polynomial Regression Analysis

Dataset	Function
DDSM	$a=-1.587629953 \times 10^{-2} \times n^2 + 1.547986999 \times 10^{-1} \times n + 79.82002872$
UCI	$a=-1.330346922 \times 10^{-4} \times n^2 + 1.264906673 \times 10^{-2} \times n + 87.71053348$

In the above regression a represents accuracy and n represents the number of neurons. In Table V below the vertex has been calculated using equation 4 with the result being rounded to a whole number. The Lower and Upper bounds for the proposed search space have then been calculated and tabulated.

TABLE V. CALCULATED VERTEX, UPPER AND LOWER SEARCH BOUNDS FOR CLASIFICATION ACCURACY

Dataset	Vertex	Lower	Upper
DDSM	49.00	34.00	64.00
UCI	48.00	33.00	63.00

Following determination of the search paramaters an evaluation of the highest performing classification accuracy was undertaken. The results below in Table VI show the highest classification accuracy obtained for each dataset.

TABLE VI. HIGHEST PERFORMING CLASSIFIER PERFORMANCE

Dataset	Neurons	Accuracy [%]
DDSM	60	86.00
UCI	41	89.17

The configuration that yielded the highest accuracy existed within the predicted upper and lower bound ranges for both datasets.

The advantage of this technique is that a smaller subset of baseline experiments were able to be performed that reduced the computational resources associated with training. The predicted upper and lower bounds represented the range in which the best performing classifier was found. While some may argue that the regression did not predict the highest yielding classifier there should be no expectation that this is the case otherwise the problem domain should be too simplistic and would warrant more efficient techniques rather than necessitating a pattern recognition approach.

An analysis of sample variance was performed across the dataset (Table VII) and also between the upper and lower bounds.

TABLE VII. SAMPLE VARIANCE ANALYSIS

Dataset	Full Dataset	Bounded Range
DDSM	2.2388111	1.7246755
UCI	0.5200425	0.4079255

The calculation of the sample variance indicated that the bounded ranges were less variable than the full population. The higher value in the DDSM dataset for population variance infers that the DDSM dataset is harder to classify or varies more than the UCI dataset. This could infer that the DDSM dataset is more complex to model than the UCI dataset. Despite potentially the greater complexity in the DDSM dataset a polynomial regression was still able to predict an upper and lower bound that reduced search effort and provided a degree of confidence that the best configuration had been achieved.

From the results obtained it appears that the technique works on breast cancer datasets of varying degrees of complexity. The advantage of the technique is that it can reduce the expense of classifier training while providing some confidence that a more accurate classifier has been selected.

This technique may have the disadvantage of requiring baseline experiments to be run as part of determining the search bounds however this also ensures its applicability to the algorithm and dataset utilised.

The advantage is that for large datasets we can potentially be evaluating if a system can be achieving higher degrees of accuracy within time and resource constraints. This allows effective use on the big datasets that are becoming the organisational reality.

A comparison with techniques that have been proposed to predict either the best performing number of neurons or an upper and lower bound appears in Table VIII below. It is noted that none of the techniques below identified the best performing topology.

TABLE VIII. COMPARISON TO PREDICTED NUMBER OF NEURONS USING OTHER TECHNIQUES

Author	Neurons	
	DDSM	Wisconsin
Arai [11]	21	341
Boger & Guterman [19]	4	7
Berry & Linoff [20]	12	20
Blum [21]	2-6	3-10
Proposed technique	34-64	33-63

VI. CONCLUSIONS

We have proposed and investigated a novel technique for predicting the best number of neurons in neural network classifiers. An accuracy of 86% (60 neurons) was achieved on the DDSM dataset and 89.17% (41 neurons)was achieved on the UCI dataset. The classification accuracies were predicted within the lower and upper bounds estimated for each dataset. Thus polynomial regression was a good predictive technique on the two datasets.

Our future research will investigate utilising the technique on large datasets to determine if the predictive capacity holds true.

REFERENCES

[1] J. Schmidhuber, “Deep learning in neural networks: An overview”, Neural Networks, 2015, vol. 61, pp. 85-117.

[2] J. Liu, J. Chen, X. Liu and J. Tang, “An Investigate of Mass Diagnosis in Mammogram with Random Forest,” *in Fourth International Workshop on Advanced Computational Intelligence* (Wuhan, Hubei, China: IEEE, 2011), pp. 638-641.

[3] American Cancer Society, Cancer Facts & Figures 2014, Atlanta: American Cancer Soeciety, 2014.

[4] D. Roses, Clinical Assessment of breast cancer and benign breast disease, Chapter 14, In M. Harris (Ed), Breast Cancer, Churchill Livingstone, Philadelphia, 2005, pp. 15-26.

[5] Food and Drug Administration, MQSA National Statistics, available at http://fda.gov/Radiation-EmittingProducts/MammographyQualityStandasActandProgram/FacilityScorecard/ucm113858.htm accessed 8 December 2014.

[6] T. Wilk and M. Wozniak, "Soft computing methods applied to combination of one-class classifiers", Neurocomputing, vol 75, pp. 185-193.

[7] T. Vogl, J. Mangis, A. Rigler, W. Zink and D. Alkon, "Accelerating the convergence of the backpropagation method", Biological Cybernetics, 1988, vol. 59, pp. 257-263.

[8] G. Huang, "Learning capability and storage capacity of two-hidden layer feedforward networks", IEEE Transactions on Neural Networks, 2003, vol. 14, no. 2, pp. 274-281.

[9] S. Huang and Y. Huang, "Bounds on number of hidden neurons of multilayer perceptron's in classification and recognition", in IEEE International Symposium on Circuits & Systems, 1990, vol. 4, pp. 2500-2503.

[10] S. Tamura and M. Tateishi, "Capabilities of a four-layered feedforward neural network: Four layers versus three", IEEE Transactions on Neural Networks, 1997, vol. 8, no. 2, pp.251-255.

[11] M. Arai, "Bounds on the number of hidden units in binary-valued three-layer neural networks", Neural Networks, 2008, vol. 6, pp. 855-860, 1993.

[12] H. Yuan, F. Xiong and X. Huai, "A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy", Computers and Electronics in Agriculture, 2003, vol. 40, pp. 57-64.

[13] Z. Zhang, X. Ma and Y. Yang, "Bounds on the number of hidden neurons in three-layer neural networks", 2003, vol. 16, pp. 995-1002.

[14] E.D. Ubeyli, "Implementing automated diagnostic system for breast cancer detection", Expert Systems with Applications, 2007, vol. 33, no. 4, pp. 1054-1062.

[15] S. Bashir, U. Qamar and F.H. Khan, "Heterogeneous classifiers fusion for dynamic breast cancer diagnosis using weighted vote based ensemble", 2014, Quality and Quantity, doi: 10.1007/s11135-014-0090-z.

[16] M. Heath, K. Bowyer, K. Kopans, D. Moore and P. Kegelmeyer, "The digital database for screening mammography", Medical Physics Publishing, IWDM-2000, 2001.

[17] W. Wolberg, O. Mangasarian, D. Aha, "UCI machine learning repository", University of Wisconsin Hospital, 1992, available at http://www.ics.uci.edu/mlearnIMLRepository.html.

[18] P. Pujari and Y. Gupta, "Improving classification accuracy using feature selection and ensemble model", International Journal of Soft Computing and Engineering, 2012, vol. 2, no. 2, pp. 380-386.

[19] Z. Boger and H. Guterman, "Knowledge extraction from artificial neural network models", IEEE Conference on Systems, Man, Cybernetics and Simulation, 1997, vol. 4, pp. 3030-3035.

[20] M.J.A. Berry and G. Linoff, *Data Mining Techniques*, NY: John Wiley & Sons, 1997.

[21] A. Blum, *Neural Networks in C++*, JY: John Wiley & Sons, 1992