A statistical comparison between an unsupervised neural network and a partially connected neural network in the detection of breast cancer

Smaranda Belciug

Abstract. This paper deals with the comparison of the two neural network methods of learning: supervised (partially connected neural network) and unsupervised (self organizing feature maps (SOFM), in order to assess their performances on a labeled breast cancer database. A statistical comparison has been made to reveal the differences between the two methods regarding diagnosis accuracy and computational time.

1. Introduction

Breast cancer is the second most common type of cancer (10.4% of all cancer incidence, both sexes counted - World Health Organization/International Agency for Research on Cancer, June 2003), after lung cancer, and the fifth most common cause of cancer death. In 2004, breast cancer caused 7% of cancer deaths, and almost 1% of all deaths -World Health Organization, February 2006. According to the American Cancer Society (2009), the chance of a woman having invasive breast cancer some time during her life is a little less 1 in 8, and the chance of dying from breast cancer is about 1 in 35. Luckily, breast cancer death rates have been going down due to better treatment and earlier detection. Breast cancer detection diagnosis is achieved either by tissue biopsy or by using conventional imaging (CI), or the more expensive nuclear imagining, magnetic resonance imaging (MRI), or positron emission tomography (PET) etc. The average accuracy of using such modern medical imaging methods for detecting breast cancer or recurrent events ranges from 75% to 95% (see Vranjesevic et al., 2002 [13]; Gilbert et al., 2008[5]; Lee et al., 2009 [11]). Hence, developing an intelligent system to predict initial breast cancer would assist health professionals in making informed decision.
The use of NN in detecting breast cancer is widespread, because providing a specific algorithm on how to identify the disease is not needed. Thus, Fogel et al.1997 trained a NN using evolutionary programming for the detection of breast cancer using radiographic features and patient age. Revett et al. (2005) [12] and Gorunescu et al. (2007) [4] designed a medical decision support system for breast cancer based on a hybrid model containing rough sets and probabilistic neural networks. Kuo et al. (2008 [10]) used NN as a classifier for the evaluation of the (3D) power Doppler ultrasound in the differential diagnosis of solid breast tumors. Hsiao et al. (2009) [7] trained a MLP classifier using the vascularity indices (harmonic and non-harmonic

3D power Doppler imaging) for determining whether the breast tumors are benign or malign.

This paper presents a statistical comparison between two types of neural networks: an unsupervised neural network, *self organizing map network* (SOM), and a new version of *partially connected supervised neural network* (PCNN), introduced by Belciug et al. 2010 [2].

The remaining paper is organized as follows: Section 2 describes the SOM algorithm, Section 3 presents an overview of the PCNN, detailing the process of inhibiting the synapses, Section 4 describes the breast cancer database. In Section 5 the computational results are discussed and evaluated. The paper ends with conclusions in Section 6.

2. SOM

A self organizing (feature) map (SOM) is an unsupervised neural network which can be used for clustering tasks. In a SOM there are two layers of neurons: an input layer and a competition layer. Being an artificial NN, a SOM has two phases: training and testing (mapping). In the training phase the map is being built using input samples. In the mapping phase the SOM automatically classifies a testing sample.

Technically, a SOM consists of nodes (neurons). Each node is associated with a weight vector, having the same dimension as the input samples, and has a position in the map space. All the nodes are interconnected with each other, thus allowing lateral excitation or inhibition. A sample vector is placed onto the map by finding the node with the closest weight vector to the sample vector and by assigning the map coordinates of this node to the sample vector.

Basically, for each training sample, the Euclidian distance is computed between it and all the nodes weights. The node which has assigned the minimum value of the Euclidian distance is the *best matching unit* (BMU) for that sample. It and a number of neurons that are in its neighborhood are updated using the following formula:

$$w_{neuron}(t+1) = w_{neuron}(t) + \theta_{neuron(t)} \cdot \alpha(t) \cdot D(t) - w_{neuron}(t)) \qquad (1)$$

where $w_{neuron}(t+1)$ is the updated value of a certain weight w, the neighborhood function:

$$\theta(t) = \begin{cases} 1 & \text{if } neuron \in neighborhood(BMU) \\ 0 & \text{if } neuron \notin neighborhood(BMU) \end{cases} \qquad (2)$$

shrinks with time, $D(t)$ is the input vector and $\alpha(t)$ - monotonically decreasing learning coefficient.

Initially, the neighborhood is broad. When the neighborhood has shrunk enough (just a couple of neurons) the weights converge to the BMUs. For each BMU the number of times it had been selected is computed. At the end of the training process there will be selected only a number of BMUs, the BMUs which have been selected most of the times will represent the centroids of the clusters. The number of clusters is given by the number of BMUs.

3. PCNN

The PCNN network is in fact a multi-layered feed-forward network (MLP) with some deactivated synaptic connections. The PCNN consists of a number of n of input units, representing the number of (predictive) attributes in each dataset. Important issues in the MLP design include specification of the number of hidden layers and the number of units in these layers (Haykin, 1990 [6]). In this study we have chosen heuristically both the number of hidden layers and the number of hidden units in order to obtain good prediction accuracy, and, at the same time, a simplified network. Consequently, we used a network with two hidden layers. The first hidden layer contains three neurons, and the second hidden layer contains two units (one for each decision class A and B), and one output unit, representing the class attribute A and B.

The network output is computed using the *winner-takes-all rule* (i.e. the highest activation unit gives the class). Technically, for each training pattern presented to the network and for each weight vector, the values corresponding to the two hidden layers are computed. According to the winner-takes-all rule, if the maximum value is given by the hidden neuron A, then the decision is A, otherwise the decision is B.

The training of PCNN is accomplished by using the backpropagation (BP) algorithm. At the beginning of the training phase, we used a fully connected neural network. The training data set consists of input signals, assigned with corresponding target (desired output, *class A* or *class B*). The network training is an iterative process and at each iteration the weights are modified using new samples from the training dataset. As for an activation function, we have used the classical sigmoid function $f(x) = \frac{1}{1+e^{-x}}$. For all units, the activity is propagated forward by $x'_i = f_i(\sum_{j \in A_i} w_{ij} x_j)$.

The signal propagates through the output layer. The output signal of the network is then compared with the desired output value (the target), which is found in the training dataset. The difference is the error signal δ_{output} of the output layer unit. The error signal δ_{output} (computed in the teaching step) is then propagated back to all the neurons, which output signals were input for the discussed neuron. Each neuron has an error signal, all of them being contained in a vector $\delta = (\delta_1, \delta_2, ..., \delta_n, \delta'_1, \delta'_2, \delta'_3)$, ($\delta_i$ is computed in the teaching step, using the following formula: $\delta_j = \sum_{i \in \delta} w_{ij} \delta_i$). The direction of data flow is changed as signals are propagated from output to input, one after the other. When the error signal for each neuron is computed, the weights of each input neuron can be modified using the following equation:

$$w^*_{(x_i)j} = w^*_{(x_i)j} + \eta \delta_i \cdot \frac{df(\sum_{j \in A_i} x_i \cdot wij)}{d(\sum_{j \in A_i} x_i \cdot wij)} x_i \tag{3}$$

After a certain number of training samples presented to the network, the weights that did not suffer major modifications, i.e. did not surpass a certain threshold, throughout the BP process, are erased from the networks architecture, being inhibited, thus their weights are set to 0 (Belciug et al. 2010 [2]).

Java implementation

A Java implementation of the above PCNN system has been carried out. What is

so important about the Java implementation of the program is that all data about patients collected by health workers can, at any time, be added, modified or deleted, with no change in the source of the program. That is because we have used JDBC *(Java Database Connectivity)* for the processing of the data. It is also worth noting that the structure of the database can be modified, for instance by adding new parameters, without affecting the functionality of the program.

4. Breast cancer database

We applied both the unsupervised neural network and the partially connected neural network on the Wisconsin Prognostic Breast Cancer -WPBC, consisting of 683 cases with two decision classes: *benign* 444 (65%) instances and *malign* 239 (35%) instances. The database contains nine ordinal (categorical) attributes (detailed description of the WPBC database at UCI Machine Learning repository: http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/). Below there are presented two samples from the database illustrating both the benign (patient #1) and malign (patient #2) cases.

TABLE 1. Example of benign/malign samples

Attributes	P1	P2
Diagnosis	*benign*	*malign*
Clump Thickness	5	8
Uniformity of Cell Size	1	10
Uniformity of Cell Shape	1	10
Marginal Adhesion	1	8
Single Epithelial Cell Size	2	7
Bare Nuclei	1	10
Bland Chromatin	3	9
Normal Nucleoli	1	7
Mitoses	1	1

Next, some words regarding each attribute:
- The first predictive attribute is "clump thickness", which represents the thickness of the group of cells that clump together forming a benign or a malign tumor. The larger the number which represents the value of this feature, the thicker the clump.
- The next feature is the "uniformity of the cell size". In the classification of breast cancer the uniformity of the size of the tumor cells has an important role. It tells us whether or not the tumor has spread to the lymph nodes, in the armpits, and whether the tumor has metastazed or spread to other parts of the body (the larger the value, the worse the prognosis).
- The "uniformity of cell shape" represents the next attribute. When the cells become differentiated they take different shapes to function as part of an organ. Cancer cells lose that ability. The cells that normally line up in the breast ducts now become disorganized. Cell division becomes uncontrolled and the cell nuclei become less uniform (the higher the deformity, the worse the prognosis).

- The "marginal adhesion" is a feature that is critical to tumor cell growth and metastasis (the larger the value, the worse the prognosis). -Most breast cancers are derived from the "epithelium" lining the ducts or lobules. The benign tumor is the proliferation of cells within the epithelial tissue, and the malign tumor invades the surrounding tissue, the larger the number, the faster the invasion.
- The presence of "bare nuclei" indicates the benignity of the cell; otherwise the absence of the bare nuclei favors the diagnosis of malignancy.
- The "bland Chromatin" is the complex combination of DNA and protein that makes up chromosomes. The function of chromatin influences the mitosis(the smaller the value, the better the prognosis).
- The "Irregularly shaped nucleoli" are characteristic for malignant cells, thus the value of the attribute should be smaller.
- The last feature of the database refers to "mitoses". Cancer is essentially a disease of mitoses. The normal "checkpoints" that supervise the mitoses are ignored by the cancer cell. Cancer begins when a single cell is transformed from a normal one into a cancer cell during the process of mitoses.

5. Results

Since the main goal of this paper is to compare the diagnosing performances of the two NN models, an *a priori* statistical power analysis (*two-tailed* type of null hypothesis) has been performed to determine the appropriate sample size in order to achieve adequate statistical power (Altman, 1991 [1]). Accordingly, a sample of 100 different computer runs for each NN model is considered, providing a statistical power equalling 99% with type I error $\alpha = 0.05$ for the comparison tests subsequently used. As a verification method, the 10-fold cross-validation is used (Kohavi, 1995 [8]). Hence, the classification accuracy training /testing is computed 10 times, each time leaving out one of the sub-samples from the computation and using that sub-sample as a test sample for cross-validation, so that each sub-sample is used 9 times as training sample and just once as testing sample (*complete cross-validation cycle*). The NN correct classification rates, computed for each of the 10 runs of the NN model, are then averaged to give the 10-fold estimate of the classification accuracy. Technically, each NN model is run 100 times in a complete cross-validation cycle, and both the training and the testing correct classification rates (i.e. training/testing performance) were recorded. Finally, eight samples, two for each NN model (one for training and other for testing performance), with equal size (100), were selected for subsequent statistical analysis.

The results of the diagnosing performances of the two NN models, in terms of mean and standard deviation (SD), averaged over 100 computer runs of a complete cross-validation cycle, are displayed in Table 2. It is worth mentioning that the distribution of the mean performance is nearly Gaussian, whatever the distribution of the values, since the samples size equals 100 (Altman, 1991 [1]).

TABLE 2. NN average training/testing performances over 100 runs

NN type	Training performance mean/SD (%)	Testing performance mean/SD (%)
SOM	88.92 / 4.12	87.64 / 4.36
PCNN	95.51 / 2.82	94.21 / 3.25

From the above table, it can be seen that:

- There is no significant difference between the training and the testing performance regardless of the neural network model used, denoting the fact that the models cross-validates well. Since the two values are close enough, we may be reasonably confident that the networks will generalize successfully. To prove this hypothesis we have used two-sided t-test for difference between two means with threshold p-value $= 0.05$ (Altman, 1991). Thus, for SOM we obtained $p = 0.51$, while for PCNN $p = 0.35$, both larger than the standard threshold 0.05, meaning that there is no statistically significant difference between training and testing accuracies.
- The diagnosis accuracy of both the SOM and PCNN is in accordance with the reported modern medical imaging tests (ranging from 75% to 95%), and almost the same performance has been obtained using this intelligent system.
- The PCNN had performed better than SOM due to the fact that the PCNN uses supervised learning, while SOFM uses unsupervised learning.

The benchmarking process for the CPU time involved 100 neural networks, from which we retained the best 50, for both the unsupervised neural network and the partially connected neural network. The results are presented in Table 3 below.

TABLE 3. Comparison Between PCNN and SOM CPU time

Database	Average CPU time SOM	Average CPU time PCNN
WPBC	1157	1046

As expected, the PCNN speed slightly exceeds the SOM speed, It is expected that for larger databases, the difference will become more significant, showing the effectiveness of such simplified NN architecture.

6. Conclusions

Both the diagnosis accuracies of SOM and PCNN were in accordance with the reported modern medical imaging experience, but much cheaper and faster. Even if in this study the results provided by the PCNN method were far better, in diagnosis accuracy and computational time, than the results of the SOM, one must recall that in real life situations we do not always have labeled database, which is needed in supervised learning (PCNN), and thus only an unsupervised neural network (SOM) will be able to make a good decision without the benefit of a supporting teacher .

References

[1] D.G. Altman, *Practical statistics for medical research*, Chapman and Hall, 1991.

[2] S. Belciug and E. El-Darzi, A partially connected neural network-based approach with application to breast cancer detection and recurrence, *IEEE Conference on Intelligent Systems*, IS2010, London, UK, (2010).

[3] D.B. Fogel, E.C. Wasson, E.M. Boughton and V.W. Porto, A step toward computer-assisted mammography using evolutionary programming and neural networks, *Cancer Lett.* **119** (1997), no. 1, 93–97.

[4] M. Gorunescu, F. Gorunesc and K. Revett, Investigating a Breast Cancer Dataset Using a Combined Approach: Probabilistic Neural Networks and Rough Sets, *Proc. 3rd ACM International Conference on Intelligent Computing and Information Systems -ICICIS07*, Cairo, Egypt, Police Press, (2007), 246–249.

[5] F.J. Gilbert, S.M. Astley, M.G.C. Gillian, O.F. Agbaje, M.G. Wallis, J. James, C.R.M. Boggis and S.W. Duffy, Single Reading with Computer-Aided Detection for Screening Mammography, *N. Engl. J. Med.* **359** (2008), no. 16, 1675–1684.

[6] S. Haykin, *Neural Networks. A comprehensive foundation*, Prentice Hall International. Inc., 1999.

[7] Y.H. Hsiao, Y.L. Huang, W.M. Liang, S.J. Kuo and D.R. Chen, Characterization of benign and malignant solid breast masses: harmonic versus nonharmonic 3D power Doppler imaging, *Ultrasound Med. Biol.* **35** (2009), no. 3, 353–359.

[8] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, *Fourteenth International Joint Conference on Artificial Intelligence*, **2** (1995), no. 12, 1137–1143.

[9] T. Kohonen, S*elf-Organizing Maps. Third*, extended edition. Springer, 2001.

[10] S.J. Kuo, Y.H. Hsiao, Y.L. Huang and D.R. Chen, Classification of benign and malignant breast tumors using neural networks and three-dimensional power Doppler ultrasound, *Ultrasound Obstet. Gynecol.* **32** (2008), no. 1, 97–102.

[11] J.H. Lee, E.L. Rosen and D.A. Mankoff, The Role of Radiotracer Imaging in the Diagnosis and Management of Patients with Breast Cancer: Part 1 Overview, Detection, and Staging, *Journal of Nuclear Medicine* **50** (2009), no. 4, 569–581.

[12] K. Revett, F. Gorunescu, M. Gorunescu, E. El-Darzi and M. Ene, A Hybrid Breast Cancer Diagnosis System: A Combined Approach Using Rough Sets and Probabilistic Neural Networks, *Proc. IEEE International Conference on Computer as a tool-Eurocon2005*, Belgrade, Serbia, (2005), 1124–1127.

[13] D. Vranjesevic, J.E. Filmont, J. Meta, D.H. Silverman, M.E. Phelps, J. Rao, P.E. Valk and J. Czernin, Whole-Body 18F-FDG PET and Conventional Imaging for Predicting Outcome in Previously Treated Breast Cancer Patients, *Journal of Nuclear Medicine* **43** (2002), no. 3, 325–329.

(Smaranda Belciug) Department of Computer Science, University of Craiova, 13 A.I. Cuza Street, Craiova, 200585, Romania
E-mail address: `smaranda.belciug@inf.ucv.ro`