ISTANBUL UNIVERSITY –
JOURNAL OF ELECTRICAL & ELECTRONICS ENGINEERING

YEAR : 2004 (1149-1153)
VOLUME : 4
NUMBER : 2

BREAST CANCER DIAGNOSIS USING STATISTICAL NEURAL NETWORKS

Tüba KIYAN[1] **Tülay YILDIRIM[2]**

[1,2] Electronics and Communication Eng. Department
Yildiz Technical University
Besiktas, Istanbul 34349 TURKEY

[1]E-mail: tkiyan@yildiz.edu.tr [2]E-mail: tulay@yildiz.edu.tr

ABSTRACT

Breast cancer is the second largest cause of cancer deaths among women. The performance of the statistical neural network structures, radial basis network (RBF), general regression neural network (GRNN) and probabilistic neural network (PNN) are examined on the Wisconsin breast cancer data (WBCD) in this paper. This is a well-used database in machine learning, neural network and signal processing. Statistical neural networks are used to increase the accuracy and objectivity of breast cancer diagnosis.

Keywords: *Radial basis function, general regression neural networks, probabilistic neural network, wisconsin breast cancer data.*

1. INTRODUCTION

The automatic diagnosis of breast cancer is an important, real-world medical problem. 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 computerized diagnostic tools, intended to aid the physician in making sense out of the confusion of data [1].

There is much research on medical diagnosis of breast cancer with WBCD data in neural network literature. In [2], a learning algorithm that combines logarithmic simulated annealing with the Perceptron algorithm is used and reported accuracy is 98.8%. In [3], the classification technique uses fuzzy modeling and cooperative coevolution reaching to a classification accuracy result of 98.98% over the entire WBCD database. In [4], the classification is based on a Feed forward Neural Network Rule Extraction Algorithm. The reported accuracy is 98.24%. The first 367 chronologically collected instances where used in [6], where the reported accuracy is 93.7%.

This breast cancer database was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. The database contains 699 samples with 683 complete data and 16 samples with missing attributes. There are

Received Date : 05.03.2003
Accepted Date: 15.06.2004

9 integer-valued attributes and each data values range from 1 to 10, as follows:
(1) Lump Thickness;
(2) Uniformity of Cell Size;
(3) Uniformity of Cell Shape;
(4) Marginal Adhesion – fibrous bands tissue that form between two surfaces;
(5) Single Epithelial Cell Size – the size of a single cell that forms tissues that lines the outside of the body and the passageways that lead to or from the surface;
(6) Bare Nuclei;
(7) Bland Chromatin–evaluates for the presence of Barr bodies;
(8) Normal Nucleoli;
(9) Mitoses – cell growth.
These attributes measure the external appearance and internal chromosome changes in nine different scales. There are two values in the class variable of breast cancer: benign (non-cancerous) and malignant (cancerous), which is represented numerically by 2 and 4 respectively [6]. Table 1 shows the data distributions.

Table 1. Classes And Their Data Distributions

Class	Total Data Number	Number of Training Data	Number of Test Data
2	444	222	222
4	239	120	119

This paper compares the statistical neural networks with multi layer perceptron on the WBCD database. Radial Basis Function (RBF), General Regression Neural Network (GRNN) and Probabilistic Neural Network (PNN) were used to classify WBCD data and these results were compared with Multilayer Perceptron (MLP).

2. APPLIED NEURAL NETWORK STRUCTURES

2.1. Radial Basis Functions (RBF) [7]:

RBF is a different approach by viewing the design of a neural network as a curve-fitting problem in a high-dimensional space. According to this viewpoint, learning is equivalent to finding a surface in a multidimensional space that provides a best fit to the training data, with the criterion for “best fit” being measured in some statistical sense. The construction of a radial-basis function network in its most basic form involves three entirely different layers. The input layer is made up of source nodes. The second layer is a hidden layer of high enough dimension, which serves a different purpose from that in a multilayer perceptron. The output layer supplies the response of the network to the activation patterns applied to the input layer. The transformation from the input space to the hidden-unit space is nonlinear where as the transformation from the hidden-unit space to the output space is linear.

2.2. Probabilistic Neural Networks (PNN) [8]:

The PNN introduced by Specht is essentially based on the well-known Bayesian classifier technique commonly used in many classical pattern-recognition problems. Consider a pattern vector x with m dimensions that belongs to one of two categories K_1 and K_2. Let $F_1(x)$ and $F_2(x)$ be the probability density functions (pdf) for the classification categories K_1 and K_2, respectively. From Bayes’ discriminant decision rule, x belongs to K_1 if

$$\frac{F_1(x)}{F_2(x)} > \frac{L_1}{L_2}\frac{P_2}{P_1} \tag{1}$$

Conversely, x belongs to K_2 if

$$\frac{F_1(x)}{F_2(x)} < \frac{L_1}{L_2}\frac{P_2}{P_1} \tag{2}$$

where L_1 is the loss or cost function associated with misclassifying the vector as belonging to category K_1 while it belongs to category K_2, L_2 is the loss function associated with misclassifying the vector as belonging to category K_2 while it belongs to category K_1, P_1 is the prior probability of occurrence of category K_1, and P_2 is the prior probability of occurrence of category K_2. In many situations, the loss functions and the prior probabilities can be considered equal. Hence the key to using the decision rules given by equations (1) and (2) is to estimate the probability density functions from the training patterns.

In the PNN, a nonparametric estimation technique known as Parzen windows is used to construct the class-dependent probability density functions (pdf) for each classification category required by Bayes' theory. This allows determination of the chance a given vector pattern lies within a given category. Combining this with the relative frequency of each category, the PNN selects the most likely category for the given pattern vector. Both Bayes' theory and Parzen windows are theoretically well established, have been in use for decades in many engineering applications, and are treated at length in a variety of statistical textbooks. If the jth training pattern for category K_1 is x_j, then the Parzen estimate of the pdf for category K_1 is

$$F_1(x) = \frac{1}{(2\pi)^{m/2}\sigma^m n}\sum \exp\left[-\frac{(x-xj)^T(x-xj)}{2\sigma^2}\right] \quad (3)$$

where n is the number of training patterns, m is the input space dimension, j is the pattern number, and σ is an adjustable smoothing parameter.

However, the choice of σ in general has been found to be not too sensitive to variations in its value.

2.3. Generalized Regression Neural Networks (GRNN) [9]:

The generalized regression neural networks (GRNNs) are the paradigms of radial basis function (RBF) networks, often used for function approximations. It's another term for Nadaraya-Watson kernel regression, and has the following form for the function mapping.

$$y(x) = \frac{\sum_k t_k \exp\left\{-\|x-x_k\|^2/2h^2\right\}}{\sum_k \exp\left\{-\|x-x_k\|^2/2h^2\right\}} \quad (4)$$

GRNNs share a special property, namely that they do not require iterative training; the hidden-to-output weights are just the target values t_k, so the output y(x), is simply a weighted average of the target values t_k of training cases x_k close to the given input case x. It can be viewed as a normalized RBF network in which there is a hidden unit centered at every training case. These RBF units are called "kernels" and are usually probability density functions such as the Gaussians considered in (4). The only weights that need to be learned are the widths of the RBF units h. These widths (often a single width is used) are called "smoothing parameters" or "bandwidths" and are usually chosen by cross validation. GRNN is a universal approximator for smooth functions, so it should be able to solve any smooth function approximation problem given enough data. The main drawback of GRNNs is that, like kernel methods in general, they suffer seriously from the curse of dimensionality. GRNNs cannot ignore irrelevant inputs without major modifications to the basic algorithm.

3. SIMULATION RESULTS

The simulations were realized by using MATLAB 6.0 Neural Network Toolbox. Four different neural network structure, multi layer perceptron, radial basis function, probabilistic neural network and generalized regression neural network were applied to WBCD database to show the performance of statistical neural networks on breast cancer data. The spread value of RBF, PNN and GRNN was chosen 4.4, 1 and 3, respectively. In MLP, learning rate was 0.6.

3.1.Training Data Simulation:

Half of the database was used for training. 222 samples of the training data belong to benign class and 120 samples belong to malignant class. The classification results of the training set by RBF, PNN, GRNN and MLP were given in the Table 2, 3, 4, and 5.

Table 2. Classification Of Training Data By RBF

Class	Benign	Malignant
True	222	120
False	0	0

Table 3. Classification Of Training Data By PNN

Class	Benign	Malignant
True	222	120
False	0	0

Table 4. Classification Of Training Data By GRNN

Class	Benign	Malignant
True	217	113
False	5	7

Table 5. Classification Of Training Data By MLP (Average)

Class	Benign	Malignant
True	217	118
False	5	2

Table 6. Performance For Training Data Classification

Type	Performance
RBF	%100
PNN	%100
GRNN	%96.4
MLP	%98.04

RBF and PNN gives the best classification accuracy with 342 correct classifications while GRNN has the lowest accuracy with 330 correct classifications for the training set. MLP has 335 correct classifications.

3.2. Test Data Simulation :

Table 7. Classification Of Test Data By RBF

Class	Benign	Malignant
True	215	113
False	7	6

Table 8. Classification Of Test Data By PNN

Class	Benign	Malignant
True	219	112
False	3	7

Table 9. Classification Of Test Data By GRNN

Class	Benign	Malignant
True	221	116
False	1	3

Table 10. Classification Of Test Data By MLP (Average)

Class	Benign	Malignant
True	212	114
False	10	5

Table 11. Performance For Test Data Classification

Type	PERFORMANCE
RBF	%96.18
PNN	%97.0
GRNN	%98.8
MLP	%95.74

A total of 341 samples were applied to the networks as test data; that is, 50% percent of the database was used for testing. 222 samples, which belong to benign class data, and 119 samples, which belong to malignant class, were chosen for the test. The results for RBF, PNN, GRNN and MLP are shown in the Table 7, 8, 9, and 10.

For the test set GRNN gives the best classification accuracy with 337 correct classifications while MLP has the lowest accuracy with 326 correct classifications. RBF classified 328 samples correctly and PNN was the second best network with 331 correct classifications.

Overall classification performances were 96.18% for RBF, 97.0% for PNN, 98.8% for GRNN and 95.74% for MLP.

4. CONCLUSION

How statistical neural networks are used in actual clinical diagnosis of breast cancer is shown in this paper. By applying statistical neural networks, a diagnostic system that performs at an accuracy level is constructed here. In this work, the performance of statistical neural network structures was investigated for breast cancer diagnosis problem. RBF and PNN are the best classifiers in training set, however the most important result must be considered with test data since it shows the future performance of the network. GRNN gives the best classification accuracy when the test set is considered. According to overall results, it is seen that the most suitable neural network model for classifying WBCD data is GRNN. This work also indicates that statistical neural networks can be effectively used for breast cancer diagnosis to help oncologists.

REFERENCES

[1] Carlos Andres Pena-Reyes, Moshe Sipper, A fuzzy-genetic approach to breast cancer diagnosis, Artificial Intelligence in Medicine, 131–155, 1999

[2] of the LSA machine, ICONIP A. A. Albrecht, G. Lappas, S. A.
Vinterbo, C.K. Wong, L. Ohno-Machado, Two applications 2002.

[3] C. A. Pena-Reyes, M. Sipper, Fuzzy CoCo: A cooperative coevolutionary approach to fuzzy modeling, IEEE Transactions on Fuzzy Systems, 9(5):727-737, 2001.

[4] R. Setiono, Generating concise and accurate classification rules for breast cancer diagnosis, Artificial Intelligence in Medicine, 18(3): 205-217, 2000.

[5] J. Zhang, Selecting typical instances in instance-based learning, Proc. Ninth International Machine Learning Workshop, pp. 470-479, Morgan-Kaufmann, 1992.

[6] Newton Cheung, Machine Learning Techniques for Medical Analysis, Thesis for the degree of Bachelor of Engineering, 2001.

[7] S. Haykin, Neural Networks: A Comprehensive Foundation, Mac Millan College Publishing Company, 1994

[8] Anthony T. C. Goh, Probabilistic neural network for seismic liquefaction potential, NRC Research Press Web site, 2002

[9] C. Lu, J. De Brabanter, S. Van Huffel, I. Vergote, D. Timmerman, Using artificial neural networks to predict malignancy of ovarian tumors, 23rd Annual Int. Conf. Of the IEEE Engineering in Medicine and Biology Society.