ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

COMPARITIVE ANALYSIS OF ANFIS AND FRBF- SURVIVAL TIME PREDICTION OF LUNG CANCER PATIENT

S. Karthikeyeni[1], S. Ramya[2]

Associate Professor, Department of Computer Science (P.G.),

Kongu Arts and Science College, Erode[1]

Research Scholar, Kongu Arts and Science College, Erode[2]

Abstract: Lung cancer also known as carcinoma of the lung or pulmonary carcinoma is a malignant lung tumor characterized by uncontrolled cell growth in tissues of the lung. Lung cancer is the leading cancer killer in both men and women. The five-year survival rate for lung cancer is 53.5 % for cases detected when the disease is still localized (within the lungs). However, only 15 % of lung cancer cases are diagnosed at an early stage. Over half of people with lung cancer die within one year of being diagnosed. This proposes a new model for prediction of complications developing due to survival time of lung cancer patients. This system compares the performance of various neuro fuzzy techniques such as ANFIS and FRBF.

Keywords: neuro fuzzy techniques, ANFIS, FRBF.

1. INTRODUCTION

Lung cancer is a disease characterized by uncontrolled cell growth in tissues of the lung. If left untreated, this growth can spread beyond the lung in a process called metastasis into nearby tissue or other parts of the body. Most cancers that start in lung, known as primary lung cancers, are carcinomas that derive from epithelial cells. The main types of lung cancer are small-cell lung carcinoma (SCLC), also called oat cell cancer, and non-small-cell lung carcinoma (NSCLC). The most common symptoms are coughing (including coughing up blood), weight loss and shortness of breath. The most common cause of lung cancer is long-term exposure to tobacco smoke, which causes 80–90% of lung cancers. Non-smokers account for 10–15% of lung cancer cases, and these cases are often attributed to a combination of genetic factors, radon gas, asbestos, and air pollution including second-hand smoke. Lung cancer may be seen on chest radiograph and computed tomography (CT scan). The diagnosis is confirmed with a biopsy which is usually performed by bronchoscope or CT-guidance. Treatment and long-term outcomes depend on the type of cancer, the stage (degree of spread), and the person's overall health, measured by performance status. Common treatments include surgery, chemotherapy and radiotherapy. NSCLC is sometimes treated with surgery, whereas SCLC usually responds better to chemotherapy and radiotherapy. The objective of the data set is to predict the survival time of lung cancer patient based on personal data such as age, stage of cancer and the survival time of the diagnosed patient etc., try to decide the survival time of the patient. Adaptive neuro fuzzy inference system (ANFIS) is a kind of artificial neural network that is based on Takagi–Sugeno fuzzy inference system. Since it integrates both neural networks and fuzzy logic principles, it has potential to capture the benefits of both in a single framework. In this proposed work ANFIS result is compared with Radial Basis Function (RBF) based Adaptive Fuzzy System (AFS), called Fuzzy Radial Basis Function (FRBF). FRBF employs the gaussian functions to represent the membership functions of the premise part of fuzzy rules. This system is developed using MATLAB.

2. RELATED WORKS

Data mining is the extraction of hidden predictive information from large databases are powerful technology with great potential that helps to focus on the most important information in data warehouses. Modern medicine generates a great deal of information stored in the medical database. Extracting useful knowledge and providing scientific decision-making for the diagnosis and treatment of disease from the database increasingly becomes necessary. Data mining in medicine can deal with this problem. It can also improve the management level of hospital information and promote the development of telemedicine and community medicine. Because the medical information is in nature of redundancy, multi-attribution, incompletion and closely related with time, medical data mining differs from other one.

2.1 DATA MINING APPLICATIONS

The methods and applications of medical data mining are based on computation intelligence such as artificial neural network, fuzzy system, evolutionary algorithms, rough set, and support vector machine are in trends.

- An ANN often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

information using a connectionist approach to computation.

- Fuzzy sets as a means of representing and manipulating data that was not precise, but rather fuzzy. There is a strong relationship between Boolean logic and the concept of a subset, there is a similar strong relationship between fuzzy logic and fuzzy subset theory.
- SVMs are sets of supervised learning methods whose training technique permits to represent complex non linear functions. The characteristic parameters of the system are determined solving a quadratic convex optimization problem.

In the paper “Using a neuro fuzzy approach in a medical application” [3]compare Neuro-fuzzy systems conceptually and evaluate their performance, early experimental results in a medical database proved a promising performance and the need for further evaluation in other medical application.

In the paper” A Lung Cancer Mortality risk Calculator based on SEER data.” Ankit, Misra, Narayanan, Lalith and Alokchoudhary [1] ,analysed the lung cancer data available from the SEER program with the aim of developing accurate survival prediction models for lung cancer using data mining techniques.

In the paper “Conditional Survival in rectal Cancer, A SEER Database analysis” S.J.Wang, R. Emcry, C.D.Fuller [5], a few data mining applications which has become a very significant component of cancer research and survivability analysis. A number of techniques based on data mining have been proposed for the survivability analysis of various cancers.

In the paper “A Fuzzy radial basis function network ”Mitra S, Basak [4] FRBF network is designed by integrating the principles of the radial basis function network and the fuzzy c-means algorithm. The architecture of the network is suitably modified at the hidden layer to realise a novel neural implementation of the fuzzy clustering algorithm.

In the paper “Pattern recognition with fuzzy objective function algorithms.” [3] .An origin of sources of fuzziness is related to labels expressed in feature space as well as to labels of classes taken into account in classification procedures. This represents the difference between way of information processing by means of probability and fuzzy set theory and a way of interpreting the results.

3. METHODOLOGY

3.1 ADAPTIVE NEURO FUZZY INFERENCE SYSTEM

ANFIS is an adaptive network. An adaptive network is network of nodes and directional links. Associated with the network is a learning rule - for example back propagation. It’s called adaptive because some or all of the nodes have parameters which affect the output of the node. These networks are learning a relationship between inputs and outputs. The ANFIS approach learns the rules and membership functions from data. The ANFIS architecture is shown below fig 1. The circular nodes represent nodes that are fixed whereas the square nodes are nodes that have parameters to be learnt.

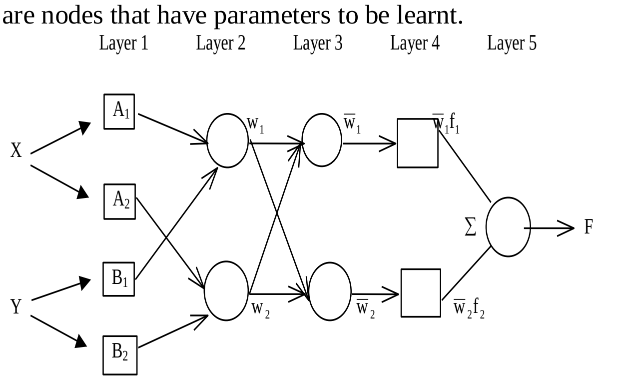
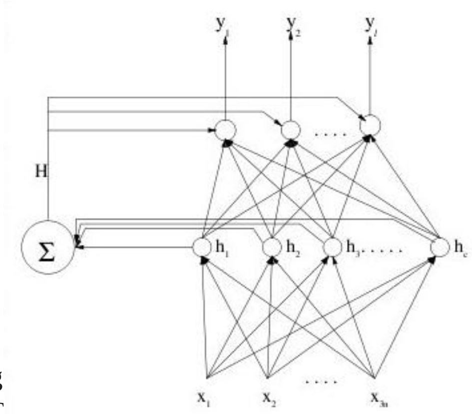


Fig 1: An ANFIS architecture

For the training of the network, there is a forward pass and a backward pass. User should look at each layer in turn for the forward pass. The forward pass propagates the input vector through the network layer by layer. In the backward pass, the error is sent back through the network in a similar manner to back propagation.

3.2 FUZZY RADIAL BASIS FUNCTION

Radial functions are a special class of function. Radial function characteristic feature is that their response decreases (or increases) monotonically with distance from a central point. The centre, the distance scale, and the precise shape of the radial function are parameters of the model, all fixed if it is linear. Radial functions are simply a class of functions. In principle radial functions could be employed in any sort of model (linear or nonlinear) and any sort of network (single layer or Multi-layer). However Broomhead and Lowe’s 1988, radial basis function networks (RBF networks) have traditionally been associated with radial functions in a single layer network such as shown in fig 2.An RBF network is nonlinear if the basis functions can move or change size or if there is more than one hidden layer.



Fig

Eac x feeds forward to m basis functions whose outputs are linearly combined with weights $\{w_j\}^{m}_{j=1}$ into the network output f(x).

4. PROPOSED WORK

4.1 ANFIS

The testing data set check the generalization capability of the resulting fuzzy inference system. The idea behind

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

using a checking data set for model validation is that after a certain point in the training, the model begins over fitting the training data set. In principle, the model error for the checking data set tends to decrease as the training takes place up to the point that over fitting begins, and then the model error for the checking data suddenly increases. Over fitting is accounted for by testing the FIS trained on the training data against the checking data, and choosing the membership function parameters to be those associated with the minimum checking error if these errors indicate model over fitting. Usually, these training and checking data sets are collected based on observations of the target system and are then stored in separate files.

```
Load dataset
For i = 1: size(dataset, 1)
F1 = iris(: , i) = 1;
F2 = iris(: ,2) = 2;
F3 =iris(: ,2) = 3;
End
[U center obj] = initfcm(cluster, size(iris, 1))
For
j = 1: size(dataset2,2)datalung = reshape(dataset2, j)
End
[center objfun] = fcm(u, 2) %% cluster Operation
For i = 1: no of iter
if i == size(dataset, 1)&& abs(objfun) < miniter
For
j = 1: size(dataset2,2)datalung = reshape(dataset(i), j)
End
break; else i < size(dataset, 2)&& mean(obj)
        < miniter
For
j = 1: size(dataset2,2)datalung = reshape(dataset(i), j)
End
lungfis = genfis1( datalung(: ,4: 5), no of inputs in dataset, memberfunction )
%% to calculate risk factor
```

Pseudo code for ANFIS

4.2 FRBF

A training set is an m labeled pair $\{Xi, di\}$ that represents associations of a given mapping or samples of a continuous multivariate function. The sum of squared error criterion function can be considered as an error function E to be minimized over the given training set (i.e.) to develop a training method that minimizes E by adaptively updating the free parameters of the RBF network. These parameters are the receptive field centers μ_j of the hidden layer Gaussian units, the receptive field widths σ_j, and the output layer weights (wij). Because of the differentiable nature of the RBF network transfer characteristics, one of the training methods considered here was a fully supervised gradient-descent method over.

```
Load dataset
For i = 1: size(dataset, 1)
F1 = iris(: , i) = 1;
F2 = iris(: ,2) = 2;
F3 =iris(: ,2) = 3;
End
[U center obj] = initfcm(cluster, size(iris, 1))
For
j = 1: size(dataset2,2)datalung = reshape(dataset2, j)
End
[center objfun] = fcm(u, 2)  %% cluster OPeration
For i = 1: no of iter
if i > 1  && abs(obj) < miniter
For  j = 1: size(dataset2,2)
datalung = reshape(dataset2, j)
End
break;
else  i == 1
For  j = 1: size(dataset2,2)          datalung =
reshape(dataset2, j)
End
end
end
lungfis = setfis(lungfis, no of inputs in dataset )
  %% to calculate risk facror in each symptoms
```

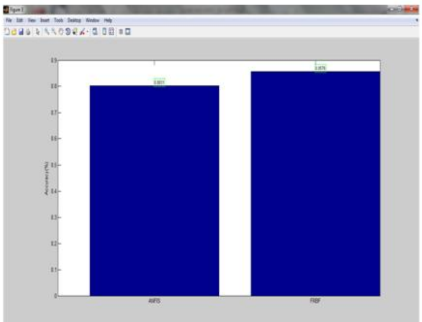
Pseudo code for FRBF

5. RESULT AND DISCUSSION

5.1 Lung Cancer Datasets

The proposed system uses the dataset of lung cancer affected patients. The data were collected from 1st January 2012 to 10th June 2014 by referring health and social care information centre's data set. The lung cancer datasets contains information such as primary tumor stages (T), regional lymph node involvement stages (N), metastases stages (M), Non–Small Cell Lung Cancer Staging, survival time of the patient, dead event status and age. The data sets are mainly based on lung cancer affected person. A total of 422 cases are available in datasets. The data sets were divided for training and testing in 60-40 ratio.

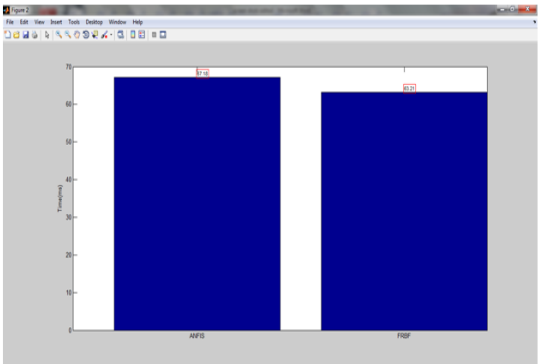
Accuracy of neuro fuzzy techniques for cancer patient



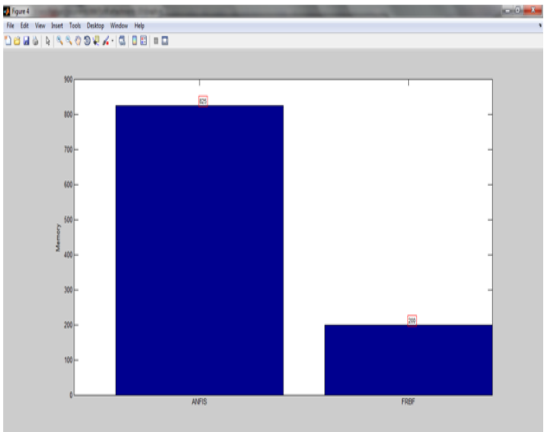
Time slot of neuro fuzzy techniques for cancer patient

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

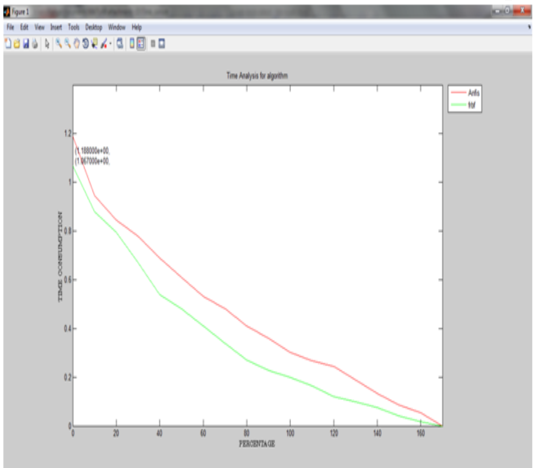
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014



Memory slot of neuro fuzzy techniques for cancer patient



Time slot chart in milliseconds of both techniques



6. CONCLUSION AND FUTURE WORKS

The system produces an accuracy level of 85% than other algorithm. The experimental result shows that the FRBF technique is superior to ANFIS technique. FRBF technique provides high level of accuracy and occupies less memory in short period of time. The system produces feasible solution for the prediction of survival time of lung cancer patient. The system can be enhanced with following features, The same method could also be applied in diagnosing other diseases like breast cancer, coronary artery, dermatological diseases etc. Increase in sample size will help or may give further analysis such as use of image to predict the stages of cancer. The soft computing techniques based lung cancer patient's living risk rate prediction system can be compared with other statistical based risk rate prediction system to evaluate the performance. The withheld continuations of the data set provided for training and measure errors can be predicted.

7. REFERENCES

1. AnkitAgrawal, Misra S, Narayanan R, Lalith and Alokchoudhary poster: A Lung Cancer Mortality risk Calculator based on SEER data. In proc. of IEEE 1st International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) pages 233-233, 2011.
2. Bezdek J C (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York.
3. Koutsojannis, Ioannis Hatzilygeroudis,(2007),Using a Neuro fuzzy Approach in a Medical Application.
4. Mitra S, BasakJ(2001) FRBF : a fuzzy radial basis function network, Neural ComputAppl 10:244-252.
5. Wang S J, Emcry R, Fuller C D , Conditional Survival in rectal Cancer, A SEER Database analysis .Gastrointest cancer Res, 1:84-89, 2007.