IJCSEC- International Journal of Computer Science and Engineering Communications.
Vol.3, Issue 2, 2015, Page.763-770, ISSN: 2347–8586

Data Analysis Classifier Model for Prediction of Human Ailments Using Artificial Neural Network

T.Rajasekaran , Assistant Professor,
Department of Computer Science and Engineering,
KPR Institute of Engineering and Technology,
Coimbatore, India
rajasekaran30@gmail.com

Abstract: Recent trends has shown that health care delivered in industrialized nations often falls short of optimal, evidence based care. Clinical data are an ever growing source of information that is generated from the hospitals in the form of patient records, when these data are mined properly, the information hidden in these data are huge resource bank for medical research. These data often contain hidden patterns and relationships, which can lead to better diagnosis, better medicines, better treatment, and overall, a platform to better understand the mechanisms governing almost all aspects of the medical domain. Classification is a machine learning technique used to solve various problems like pattern classification, image processing, etc. Neural network is the effective tool for pattern classification. Classification of any data is important to know that the data are belongs to which group. With tremendous growing population, the doctors and experts available are not in proportion with the population. Hence it focuses on computing the probability of occurrence of a particular ailment from the medical data.

1.INTRODUCTION

Given the rapidly growing population, the increased burden of chronic diseases and the increasing health care costs, there is an urgent need for the development, implementation, and deployment, in every day medical practice, of new models of health care services. In this scenario, home monitoring and data mining play an important role. Data mining is the computer assisted process of digging through and analyzing a large quantity of data in order to extract meaningful knowledge and to identify phenomena faster and better than human experts. Data Mining, also popularly known as Knowledge. Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. Data mining is widely used in business (insurance, banking, retail), science research (astronomy, medicine), and government security (detection of criminals and terrorists). Other similar terms referring to data mining are data dredging, knowledge extraction and pattern discovery to address the deficiencies in care, health care organizations are increasingly turning to clinical decision support systems, which provide clinicians with patient specific assessments or recommendations to aid clinical decision making. In recent trends, the statistics reveals that technology always stayed back when it came to diagnosis, a process that still requires a doctor's knowledge and their experience to process the sheer number of variables involved, ranging from medical history to climatic conditions, temperature, environment and various other factors.

Since the number of variables are greater than the total number of variables which no model has successfully analyzed yet. To overcome this problem, medical decision support system are becoming more and more important, which will assist doctors as well as medical students in taking correct decisions. Classification is a data mining technique used to group item based on some key characteristic. It is used to classify each item in a set of data into one of predefined set of classes or groups. It can deals with the large amount of data that are involved in processing. They are being used in different industry to easily identify the type and group to which a particular tuple belongs. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects.
There are many algorithms which are used for classification in data mining are following:

1. Decision tree induction
2. Nearest neighbor classifier
3. Artificial Neural Network

2. ARTIFICIAL NEURAL NETWORK

Artificial Neural Network is a computational model, which is based on Biological Neural Network and is often called as Neural Network (NN). An ANN is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. ANNs like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. To build artificial neural network, artificial neurons, also called as nodes, are interconnected. The architecture of NN is very important for performing a particular computation. Some neurons are arranged to take inputs from outside environment. These neurons are not connected with each other, so the arrangement of these neurons is in a layer, called as Input layer. All the neurons of input layer are producing some output, which is the input to next layer. A neural network is a set of connected input/output units in which each connection has a weight associated with it. The Artificial Neuron receives one or more inputs and sums them to produce an output. Usually the sums of each node are weighted, and the sum is passed through a function known as an activation or transfer function. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of the input tuples. Neural network learning is also referred to as connectionist learning due to the connections between units.

We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. It is shown in Fig.2.1 Typically many such input/target pairs are used, in this supervised learning, to train a network.

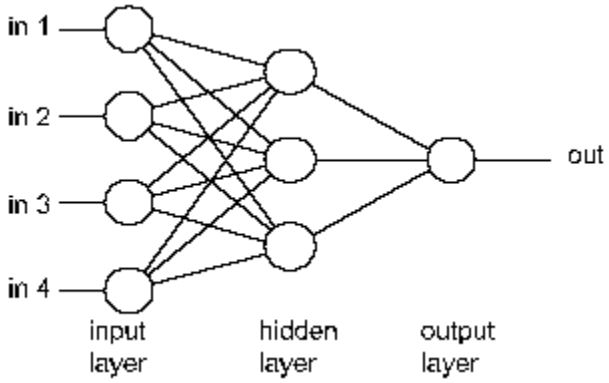

Fig 2.1 Neural Network Architecture

3. ARCHITECTURE OF NEURAL NETWORKS

The architecture of NN can be of single layer or multilayer. A single layer Neural Network can have, only one input layer and one output layer, while in multilayer neural network, there can be one or more hidden Layer.

1) Feed-Forward Networks: Feed-forward Artificial Neural Networks allow signals to travel one way only from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward Artificial Neural Networks tend to be straight forward networks that associate inputs with outputs.

2) Feedback Networks: Network with connections from a layer's output to that layer's input. Feedback networks can have signals traveling in both directions by introducing loops in the network. Feedback networks are dynamic their state is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found. Feedback architectures are also referred to as interactive or recurrent. In this paper neural network approach is proposed to identify the diseases for the particular symptoms.

4. RELATED WORKS

Many research works are done based on classification using neural network algorithm. Some of them are explained below. Makeshwar, M.S et al. proposed feed forward Back propagation algorithm was used for face deduction algorithm. In the field of image processing it is very interesting to recognize the human face for general life applications. The primary goal of face object recognition research system is to create a system which can identify human face as an object. They developed a simple architecture that recognizes the human face. They introduced a simple architecture based on add boosted classifier for face detection, simple token finding and match using back propagation neural network.

Naik, A.R et al. proposed a new technique of weather classification and forecasting using Levenberg Marquardt Back propagation feed forward neural network.It uses weather data to predict and classify the day like sunny, rainy, cloudy, etc. Weather prediction is used to estimate future weather condition. Weather condition is a state of atmosphere at given time in terms of weather variables like temperature, pressure, wind direction etc.

Weather is non linear and dynamic process it varies day-to-day even minute-to-minute.As the climatic dataset is highly non-linear, Artificial Neural Network (ANN) can be used for weather prediction and classification. Rehman. M. Z et al. proposed an algorithm for improving the current working performance of Back-propagation algorithm by adaptively changing the momentum value and at the same time keeping the gain parameter fixed for all nodes in the neural network. Three different algorithms are used for Gradient Descent with simple momentum (GDM),Gradient Descent with Adaptive Gain (GDM-AG),Gradient Descent with Adaptive momentum (GDAM).

The Gradient Descent Method with Adaptive Momentum (GDAM) works by adaptively changing the momentum and keeping the gain parameter fixed for all nodes in the neural network. The performance of the proposed GDAM is compared with Gradient Descent Method with Adaptive Gain (GDM-AG) and Gradient Descent with Simple Momentum (GDM). The performance of GDAM is verified by means of simulation on the above five classification problems. The results showed that GDAM has performed well and gave more accuracy for all classification problems than other two methods. Survival analysis is one of the classification problem. Survival analysis plays an important role not only for health care policy markers, but also for the clinician. Dhande.J.D et al developed the design of classifier using Artificial Neural Network for patients survival analysis based on echocardiography dataset. The Dataset used for this research is echocardiogram database. It is an open source and it gives the information of patients after heart attacks. They proposed the classifier designed using BPNN and RBFNN for classification of patient's survival at least one year that is death or life after a heart attack. Performance of the classifier is measured in terms of classification accuracy. Result showed that the good design of classifier model for this Echocardiogram dataset for survival analysis is Back propagation Neural Network (BPNN).The Back propagation Neural Network classifier gives 93 % classification accuracy on training set and for generalization of network test on testing set with 84 % classification accuracy. This comparison gives the BPNN artificial neural network classifier model is good classifier out performs RBFNN classifier model.

Artificial neural networks have been successfully applied to problems in pattern classification, function approximations, optimization, and associative memories.

Swain, M et al. focused on IRIS plant classification using Neural Network. This problem concerns the identification of IRIS plant species on the basis of plant attribute measurements. Classification of IRIS data set would be discovering patterns from examining petal and sepal size of the IRIS plant. In this work, Multilayer feed- forward networks are trained using back propagation learning algorithm. One of the most popular and best known databases of the neural network application is the IRIS plant dataset. The IRIS data set classifies three different classes of IRIS plant by performing pattern classification. The IRIS data set includes three classes of 50 objects each, where each class refers to a type of IRIS plant. Out of these 150 instances, 75 instances were used for training and 75 for testing. Result showed that Back propagation Algorithm gives the best accuracy. The number of epochs required to train the neural network range from 500 to 50000 and the accuracy ranges from 83.33% to 96.66%.

Neural networks have been widely used for breast cancer diagnosis. Feed forward neural networks (FFNN) are commonly used for classification. Paulin. F et al. based on classification of Breast cancer using Feed Forward Artificial Neural Networks. In this research a feed forward neural network is constructed and the Back propagation algorithm is used to train the network. The proposed algorithm is tested on a real life problem, the Wisconsin Breast Cancer Diagnosis problem.

The Wisconsin breast cancer database was used for this. In this paper six training algorithms are used. They are,Batch Gradient Descent (BGD),Batch Gradient Descent with Momentum (BGDM) Quasi Newton (QN),Resilient Back Propagation (RBP),Conjugate Gradient (CG),Levenberg Marquardt (LM).

The performance of the network is evaluated using Wisconsin breast cancer data set for various training algorithms. Among these six methods, Levenberg Marquardt method gave the good result of 99.28% accuracy. The Feed Forward Neural Network using Back Propagation Algorithm with Levenberg Marquardt training algorithm produced the best results for diagnosis.

Dr. Rani, U.K et al. proposed a paper in which Heart diseases dataset is analyzed using Neural Network approach. To perform classification task of medical data, the neural network is trained using Back propagation algorithm. Cleveland dataset concerns classification of person into normal and abnormal person regarding heart diseases. Their experiment is conducted with Heart Disease dataset by considering the single and multilayer neural network models. Back propagation algorithm with momentum and variable learning rate is used to train the networks. The experimental results proved that neural networks technique provides satisfactory results for the classification task.

Tuisima, S et al. based on the classification of game addiction level in secondary school students. They classified level of addiction using Back propagation Neural Networks and compared the results to Decision Tree Learning. This research classified computer games, based on their playing characteristics, into four categories: a) Long Term Game, b) Casual Game, c) Real Time Game, d) Turn-Based Game. Data was analyzed using two standard learning algorithms Back propagation Neural Networks and Decision Tree Learning. The results obtained from the Neural Networks yields higher percentage of accuracy than Decision Tree in Long Term Game, Turn Base Game, and Real Time Game, with average percentage of accuracy of 97.75, 97.73, and 90.0 respectively.

Gupta, A et al. presented a medical decision support system based on the neural network architecture for Medical diagnosis. The system is trained by an improved BP algorithm. They described an optimal feed forward back propagation algorithm. Feed-forward back propagation neural network is used as a classifier to distinguish between infected or non-infected person. The back propagation algorithm used for training depends on a multilayer neural network with a very small learning rate, especially when using a large training set size. It can be applied in a generic manner for any network size that uses a back propagation algorithm and achieved the best performance with the minimum epoch and training time.

4. SYSTEM ARCHITECTURE

The objective here is to develop a data classification algorithm that will be used as a general-purpose classifier. To classify any database first it is required to train the model. From literature review, Feed forward back propagation neural network algorithm works better for all classification problems when compared to other all algorithms. So the proposed training algorithm used here is a BPN. After successful training user can also give unlabeled data to classify. Classification proceeds as these steps: First create a model by running the algorithm on the training data and test the model but the training data should be normalized initially. If accuracy is low, regenerate the model, after changing features and reconsidering samples. Then identify a class label for the incoming new data.

The neural network can only handle the numeric data so it is needed to transform the character data into numeric data. In order to train neural network this data set have to be normalized. Normalization implies that all values from a set should take values in the range from 0 to 1. There are several normalization methods are available. Here input data has been normalized by the min-max normalization in the range from 0 to 1.

Where,

X = Value that should be normalized

X n = Normalized value

X min = Minimum value of the column

X max = Maximum value of the column

5. MULTILAYER FEED FORWARD NETWORK

There are many different kinds of neural networks and neural network algorithms. The most popular neural network algorithm is back propagation. A back propagation network is fully connected, layered, and feed-forward neural network. Network activation flows in one direction only from the input layer to the output layer, passing through the hidden layer. Each unit in a layer is connected in the forward direction to every unit in the next layer. Weights between units encode the network's knowledge.

The basic structure of the neural network in this paper is a multi layered feed forward neural network. The number of input units corresponds to the nature of the given problem. The number of the output units is determined by the result required. The number of hidden units depends on the accuracy obtained.

A back propagation network usually starts with a random set of connection weights. The network adjusts its weights based on some learning rules each time it sees a pair of input-output of vectors. Each pair of vectors goes through two stages of activation a forward pass and a backward pass.

The forward pass involves presenting a sample input to the network and letting activations flow until they reach the output layer. During the backward pass, the network's actual output (from the forward pass) is compared with the target output, and errors are computed for the output units. The weights connected to the output units can be adjusted to reduce those errors. The error estimates of the output units are then used to derive error estimates for the units in the hidden layers. Finally, errors are propagated back to the connections stemming from the input units.

After extensive training, the network will eventually establish the input-output relationships through the adjusted weights on the network.

6. METHODOLOGY

In this paper, a neural network tool called Neuroph Studio 2.6 is used for disease classification. It is open source software, a nice GUI neural network environment. Neuroph simplifies the development of neural networks by Providing java neural network library and GUI tool that supports creating, training and saving neural networks. It is small, well documented, easy to use, and very flexible neural network frame work. It supports like adaline, perceptron, etc. Here Neuroph Studio 2.6 was used. In order to use Neuroph we just need Java VM 1.6 installed on our computer. It supports all types of neural network architectures like adaline, perceptron, etc,

In this experiment our goal is to classify the diseases. Goal of the Error back propagation algorithm was reduced the error up to the specified level. In this experiment maximum error is 0.01. So we have to consider the total mean square error for all training attempt. In order to achieve good result, error should be less than or equal to 0.01 otherwise it is not considered as a best result.

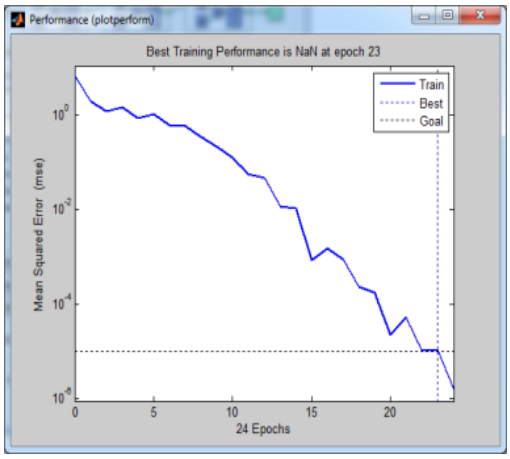

Fig 6.1 Mean Square for training attempt

7.CONCLUSION

In this system, by using Hopfield networks, Artificial Neural Network, an attempt has been made to assist the doctors to perform differential diagnosis. A pilot study was performed and the results obtained were very promising. The possibility of usage of vastly available EHR data for the purpose allows latest and continuously updated medical data available to the system. The approaches mentioned in this paper can be used to supplement and improve existing systems that provide differential diagnosis.

In the field of medical diagnosis, there is always the scope for uncertainty. This system has been built on the experience of doctors only, so there will always be a scope for ambiguous or uncertain diagnosis. It cannot be used as a substitute or a shortcut to diagnosis, but it can definitely complement the doctors' knowledge and could assist them to reach a conclusion. The doctor always has the upper hand to decide whether to use the diagnosis given by the algorithm or not. By using this system, many essential results can be obtained, reducing the dependence on human intuitions, thus reducing the effects of misdiagnosis to a great extent. After sufficient self-learning, with an extensive database of medical records to mine from, this can be used to build formidable medical assistance software that can be of great use to all doctors and specially the new practitioners and students. It will also help the medical fraternity in the long run by helping them in getting accurate diagnosis and sharing of medical practices which will facilitate faster research and save many lives.

REFERENCES

[1] Han, J., Kamber, M. Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufmann publishers.

[2] Gupta, M., Aggarwal,N. Classification techniques analysis,NCCI2010- National Conference on Computational Instrumentation, CSIO Chandigarh, India, 19-20,March 2010.

[3] Kapadia, M.T., Lakhani, A. The integration of the back propagation algorithm into an autonomous robot control system, Dwarkadas j.Sanghvi College of engineering, vile parle, Mumbai.

[4] Makeshwar, M.S., Rajgure, N.K. Supervised ANN for a Classification Problems, International Journal of Computer Science and Application Issue 2010, ISSN 0974-0767.

[5] Naik, A.R., Pathan. S.K. Weather Classification and Forecasting using Back Propagation Feed-forward Neural Network , International Journal of Scientific and Research Publications, Vol 2, Issue 12, December 2012.

[6] Rehman. M. Z., Nawi. N. M. Improving the Accuracy of Gradient Descent Back Propagation Algorithm (GDAM) on Classification Problems, International Journal on New Computer Architectures and Their Applications (IJNCAA) 1(4): 838-847, The Society of Digital Information and Wireless Communications, 2011 (ISSN: 2220-9085).

[7] Dhande.J. D., Dr.Gulhane.S.M. Design of Classifier Using Artificial Neural Network for Patients Survival Analysis, International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 1, Issue 2, November 2012

[8] Swain, M., Dash, S,K.., Dash,S and Mohapatra,A. An approach for iris plant classification using neural network, International Journal on Soft Computing (IJSC) Vol.3, No.1, February 2012

[9] Paulin.F. Santhakumaran.A. Classification of Breast cancer by comparing Back propagation training algorithms, International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 1 Jan 2011

[10] Dr. Rani, U.K. Analysis of heart diseases dataset using Neural Network Approach, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.1, No.5, September 2011

[11] Tuisima,S., Vachirapan,K and Sinthupinyo,S. Classification of Computer Game Addiction Level in Students in Secondary Education (M.1-3) using Neural Networks, 2012, 2nd International Conference on Management and Artificial Intelligence IPEDR Vol.35 (2012) © (2012) IACSIT Press, Singapore

[12] Gupta,A., Shreevastava, M. Medical Diagnosis using Back propagation Algorithm, International Journal of Emerging Technology and Advanced Engineering, Volume 1, Issue 1,November 2011