International Journal of communication and computer Technologies, ISSN: 2278-9723

Available at http://www.ijccts.org

Data Mining Based Accuracy Enhancement Of ANN Using Swarm Intelligence

Sachin Sharma[1], Dr. Savita Shiwani[2]

[1]Department of Computer Science & Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

[2] HOD, Information Technology Department, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

Received: 15-01-2014, **Revised:** 03-03-2014, **Accepted:** 09-05-2014, **Published online:** 18-06-2014

ABSTRACT

Data mining describes the process of extracting value from a database. A data-warehouse is a location and information is stored in it. The type of data stored depends largely on the type of industry and the company.

There are many tools available to a data mining practitioner. These include decision trees, various types of regression and neural networks. We use neural network and improve its accuracy. Neural networks which are used are Elman network and Feedforward network. Firstly both networks accuracy is calculated which is later optimized by use of genetic algorithm. Shortcoming of use of neural network is discussed. Neural networks are created in matlab. Optimization is performed on both the networks via Genetic algorithm which is further discussed in this paper.

Keywords—Data Mining,Neural Network,Elman Network, Feedforward Network,Swarm Intelligence,Genetic Algorithm.

1. INTRODUCTION

1.1 DATA MINING

Data mining is the process of analyzing data from different viewpoints and abridgment it into useful information, information that can be used to add to profits, cuts costs, or both. A data-warehouse is a location and information is stored in it. [1] The type of data stored depends largely on the type of industry and the company. Data mining software is one of a number of analytical tools for analyzing data.[2] It lets users to analyze data from many different extent or angles, sort it, and summarize the relationships identified. Basically, data mining is the process of discovery of correlations or patterns among dozens of fields in large relational databases. Data mining, is the computer-assisted process of digging through and analyzing massive sets of data and then extracting the meaning of the data.[3]

1.2 INTRODUCTION TO NEURAL NETWORKS

An Artificial Neural Network (ANN) is an information processing paradigm which is encouraged by the way nervous system in brain process information. The important element of this paradigm is the novel structure of the information processing system. It's composed of an excessive number of vastly interconnected processing elements (neurons) working in unity to solve specific problems. ANNs are like people they learn by example. An ANN is used for configuration of a definite application, such as pattern recognition or data classification, via learning process. Learning in biological systems involves alterations to the synaptic connections that exist linking the neurons.

Neural networks, with their remarkable capability to derive meaning from complicated or indefinite data, can be used to extract patterns and detect trends that are too complex to be noticed by other computer techniques or humans. A trained neural network can be thought of an "expert" in the category of information it is given to examine. [4]

Advantages include:

- Self-Organized: An ANN can create its own organization or representation of the information it receives during learning time.
- Adaptive learning: Ability to learn to do tasks based on the given data for training or initial experience.
- Fault Tolerance: Neural networks are well suited to work in any environment.
- Real Time Operation: ANN calculations is carried out in parallel, and special hardware devices are designed and manufactured which take benefit from this capability.

2.LIMITATION OF PREVIOUS WORK

In previous systems we organize a database with manually trained factors to obtain the output of the various systems. Neural network simulations emerge to be a recent improvement. However, this field was established before the dawn of computers.

Many significant advances have been boosted by the use of economical computer emulations. Following an early period of interest, the field survived a period of disappointment and disrepute. During early period when financial support and

professional support was least, significant advances were made by quite few researchers. The pioneers were able to build up convincing technology which surpassed the confines identified by Minsky and Papert. Minsky and Papert, published a book (in 1969) in which they summed up a general feeling of frustration (for neural networks) among researchers, and was thus accepted by most with no further analysis.

Artificial neural networks (ANNs) are increasingly used in problem domains involving classification. They are capable at finding commonalities in a set of unrelated data and for this reason are used in a growing number of classification tasks. But there lies, a commonly perceived problem with ANNs when used for classification is that, while a trained ANN can indeed classify the data, sometimes with more accuracy than a traditional, symbolic machine learning approach, the reasons for their classification cannot be found without difficulty.
Trained ANNs are commonly perceived to be dark box which map input data onto a class through a number of mathematically weighted connections between neurons. While the idea of ANNs as dark boxes may not be a problem in applications where there is little interest in the reasons behind classification, this can be a major complication in applications where it is vital to have symbolic rules or different forms of knowledge structure, such as identification which are simply interpretable by human experts. [5]

3.PROPOSED WORK

In this we enhanced the parameter used in earlier method to improve the accuracy and efficiency of the system to get the output. We made use of G.A algorithm to get the more accurate and reliable output. In general, genetic algorithms tend to work better than traditional optimization algorithms because they're less likely to be led astray/affected by local optima. The fitness factor/function is responsible for performing this evaluation and returning a positive integer number, or "fitness value", that reflects how optimal the solution is: the higher the number is, the better the solution will be.
We propose a novel, evolutionary approach which integrates traditional ANNs with genetic algorithms for extracting simple, intelligible and useful rules from trained ANNs. This approach includes advantages of artificial neural networks as well as the symbolic learning.
In brief, the proposed method makes use of a genetic algorithm to search the weight space of a trained neural network to identify the best rules for classification. The genetic algorithm uses neurons which can be mapped directly onto intelligible rules.
In these trained ANNs 70% of the systems is trained by few manually given and implemented rules to get an output. The fitness factor is manually chosen and trained to the neural networks then to get the output some set of rules are defined, accordingly to which accurate output is generated. Rests of the 30% are used for testing of these trained networks. [6]

But these trained networks are not so accurate, many a times the values generated are wrong. To increase the accuracy level of output and efficiency of the network, we proposed a new trained ANNs with genetic algorithm. Because we use the gradient descent algorithm to calculate the values of weights, neural network still encounters problems such as local minimum, slow convergence pace and convergence instability in its training process. We combine two methods to solve these problems. One solution is to improve the network algorithm. By calculation fitness factor in activation function, the speed of convergence can be increased. In addition, by compressing the weight values when they are too large, the network paralysis can be avoided.
We are taking heart data set and trying to improve the accuracy of neural network. [7]

The neural networks used are Elman neural network and Feedforward neural network.

Feed-forward ANNs they allow signals to travel one way only, from input to the output. It does not contain feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward networks are straight forward networks which connect different inputs with different outputs. They are widely used in pattern recognition. This type of association is also referred to as top-down or bottom-up approach. [8]

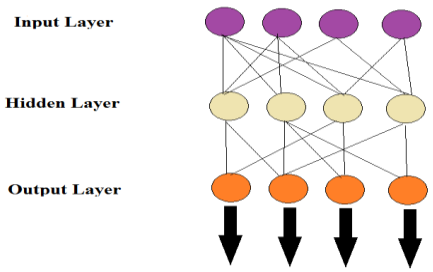

Fig.1 FeedForward Network

Elman network is a two-layer network with feedback from the first-layer output to the first-layer input. Recurrent connection allows the Elman network to detect and generate time-varying patterns.

The Elman network has purelin neurons in its output layer and tansig neurons in its hidden (recurrent) layer. This combination is special in that two-layer networks with these transfer functions can approximate any function (with a fixed number of discontinuities) with random accuracy. The only condition is that the hidden layer must have an adequate amount of neurons. More hidden neurons are required as the function being fitted increases in complexity.

Elman network differs from conventional two-layer networks in which the first layer has a recurrent connection. Delay in this connection stores values from the previous time step, those values can be used in the current time step. Thus, even if two Elman networks, with same weights and biases, are given identical inputs for a particular time step, their outputs obtained can be different because of different feedback states. As the network can store up information for future reference, it is able to learn temporal patterns and spatial patterns. The

Elman network can be trained, generate, both kinds of patterns. [9]

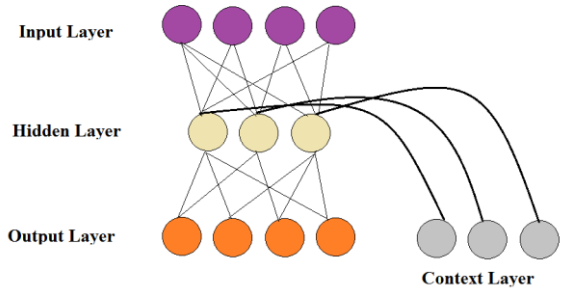

Fig.2 Elman Network

Swarm intelligence is the collective behavior of self-organized, decentralized systems artificial or natural. The concept is employed in work on artificial intelligence.

A swarm-based algorithm is proposed for the training of artificial neural networks. Training of networks like this is a difficult problem that needs an effective search algorithm to find optimal weight values. Multilayer feedforward neural networks are frequently trained by gradient-based methods, such as backpropagation, such methods may not yield a globally optimal solution. Hence to overcome the limitations of gradient-based methods, evolutionary algorithms are used to train these networks with some success. Ant colony optimization is an example of swam intelligence. [7]

Genetic Algorithms (GAs) is derived from Darwin's principal that is survival of fittest. GA is adaptive procedure. GA maintains a population of potential solutions of the candidate problem termed as individuals.
It use fitness function which is derived from the objective function of the optimization problem which is use to evaluate the individuals in a population. [10]

- Fitness function is the measure of an individual's fitness; it is used to select individuals for reproduction.
- Many of the real world problems may not have a well defined objective function and require the user to define a fitness function

Advantages of GA

- It is easy to understand and solve problem with multiple solutions.
- Fitness function: flexible expression of modeling criteria, tradeoffs amongst multiple objectives
  models optimized to definite business objectives
- Its execution technique is not based on error surface it can solve multi dimensional problems.

Table1. Accuracy of neural network(Output)

Accuracy % mapping	Normal Elman	Elman Optimized	Time(in sec)	Normal Feedforward	Feedforward Optimized	Time(in sec)
1st Iteration	94.89%	97.81%	48.61	96.35%	97.81%	250.90
2nd Iteration	95.60%	97.08%	18.25	97.81%	98.54%	30.36

4.METHODOLOGY

The efficient neural network with improved accuracy can be obtained by following given steps.

- Various parameters(data sets) are defined in initial step of any particular field such as data entries
- Select some (maximum part) among these entries for training of the neural networks and rest of the data is left for testing of the trained neural network.
- Therefore to the trained network, get the output after testing, then save the results.
- Get the weights values of the trained neural network and then update these weight values using G.A (genetic algorithm) to train the network again with these new updated values.
- In final step test the system with these new neural network optimized by genetic algorithm, it will provide the efficient and accurate results.
- If results are optimized with respect to initial network without GA then again save the results and compare them with the initial trained neural network's output.
- If the results are not optimized with respect to the initial neural network without GA then again move to update the weight values using GA
- Show the results after the comparison is completed.

5.OBJECTIVES

As per the given problem formulation main objective deals with improving efficiency of artificial neural network.

- To develop a kind of system that will optimize the system to get accuracy after testing.
- To compare results of Normal Neural Network and Optimized Neural network.
- Use Swarm Intelligence algorithm for Optimization.
- Genetic algorithm as part of Swarm Intelligence is applied.
- Developing fast and more accurate proposed algorithm.

6.CONCLUSION

According to the results accuracy of neural networks were optimized, improved significantly with the use of genetic algorithm (GA) as part of swarm intelligence. Elman network and FeedForward networks were used, their accuracy were calculated and optimized through genetic algorithm. Both normal and optimized accuracy were compared and time was calculated.

7.ACKNOWLEDGMENT

The Author sincerely thanks his parents who motivated him throughout the duration of work and Dr. Savita Shiwani (HOD (Information Technology Department) Suresh GyanVihar University, without whose guidance this work would not have been completed.

REFERENCES

[1] Berson,"Data Warehousing, Data-Mining & OLAP", TMH

[2] Agrawal, R., Imielinski, T., Swami A., Database Mining:A Performance Perspective, IEEE Transactions on Knowledge and Data Engineering, pp. 914-925, December 1993

[3] Berry, J. A., Lindoff, G., Data Mining Techniques, Wiley Computer Publishing, 1997.

[4]Jeannette Lawrence, Data Preparation for a Neural Network, Neural Network Special Report: A Miller Freeman Publication, 1992.

[5] R. Setiono, "A Neural Network Construction Algorithm which Maximizes the Likelihood Function," *Connection Science*, vol. 7, no. 2, pp. 147-166, 1995.

[6]James A .Freeman, David M. Skapura. Using neural networks and genetics algorithm to predicts stock market returns.

[7] Bhatla, N. and K. Jyoti, 2012. An analysis of heart disease prediction using different data mining techniques. Int. J. Eng. Res. Technol., 1: 1-4.

[8] R. Setiono, "A Penalty Function Approach for Pruning Feedforward Neural Networks," *Neural Computation*, vol. 9, no. 1, pp. 301-320, 1997.

[9] Gao, X.Z,; Dept. of Control Eng., Harbin INST. Of Technol,. China ;Gao, X.M,;Ovaska, S.J. A modified Elman neural network model with the application to dynamical system identification.

[10] Markus Brameier and Wolfgang Banzhaf. (2001). A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining. IEEE.p1-10

AUTHOR'S PROFILE

Sachin Sharma was born on 16th May, 1989 in Jaipur, Rajasthan. He completed his bachelor's degree (B. Tech.) in Computer Science from Suresh Gyan Vihar University, Jaipur in 2012. He is currently pursuing his master's (M. Tech.) in Software Engineering from Suresh Gyan Vihar University, Jaipur.