Applications of Neural Networks In Data Mining

M.Charles Arockiaraj

Head & Asst.Professor, Computer Science Department, Arakkonam Arts and Science College, Arakkonam-631003,

ABSTRACT:- *Finding information hidden in data is as theoretically difficult as it is practically important. With the objective of discovering unknown patterns from data, the methodologies of data mining were derived from statistics, machine learning, and artificial intelligence, and are being used successfully in application areas such as bioinformatics, banking, retail, and many others. Companies have been collecting data for decades, building massive data warehouses in which to store it.Even though this data is available, very few companies have been able to realize the actual value stored in it. The question these companies are asking is how to extract this value. The answer is Data mining.There is many technologies available to data mining practitioners, including Artificial Neural Networks, Regression, and Decision Trees. Many practitioners are wary of Neural Networks due to their black box nature, even though they have proven themselves in many situations. This paper is an overview of artificial neural networks and questions their position as a preferred tool by data mining practitioners.*

Keywords: Artificial Neural Network (ANN), neural network topology, Data mining, back propagation algorithm, Advantages.

I. INTRODUCTION:

Data mining is the term used to describe the process of extracting value from a database. A data-warehouse is a location where information is stored. The type of data stored depends largely on the type of industry and the company. Many companies store every piece of data they have collected, while others are more ruthless in what they deem to be "important". Consider the following example of a financial institution failing to utilize their data-warehouse. Income is a very important socio-economic indicator. If a bank knows a person's income, they can offer a higher credit card limit or determine if they are likely to want information on a home loan or managed investments.

Even though this financial institution had the ability to determine a customer's income in two ways, from their credit card application, or through regular direct deposits into their bank account, they did not extract and utilize this information. Another example of where this institution has failed to utilize its data-warehouse is in cross-selling insurance products (e.g. home, life and motor vehicle insurance). By using transaction information they may have the ability to determine if a customer is making payments to another insurance broker. This would enable the institution to select prospects for their insurance products. These are simple examples of what could be achieved using data mining.

II. ARTIFICIAL NEURAL NETWORKS:

An artificial neural network (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 information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.
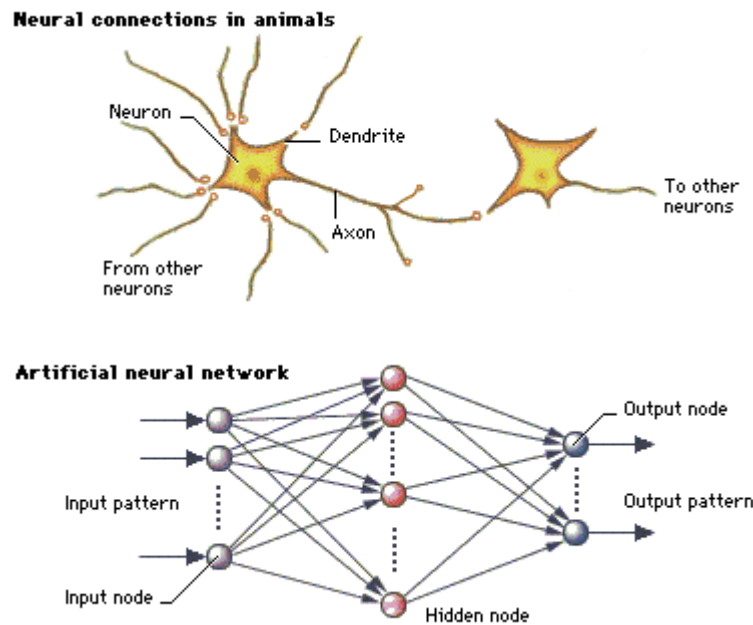
2.1 Neural Network Topologies:

Feedforward neural network: The feedforward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. The data processing can extend over multiple (layers of) units, but no feedback connections are present, that is, connections extending from outputs of units to inputs of units in the same layer or previous layers.

Recurrent network: Recurrent neural networks that do contain feedback connections. Contrary to feedforward networks, recurrent neural networks (RNs) are models with bi-directional data flow. While a feedforward network propagates data linearly from input to output, RNs also propagate data from later processing stages to earlier stages.

2.2 Training Of Artificial Neural Networks:

A neural network has to be configured such that the application of a set of inputs produces (either 'direct' or via a relaxation process) the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge. Another way is to 'train' the neural network by feeding it teaching patterns and letting it change its weights according to some learning rule. We can categorize the learning situations as follows:
• Supervised learning or Associative learning in which the network is trained by providing it with input and matching output patterns. These input-output pairs can be provided by an external teacher, or by the system which contains the neural network (self-supervised).
• Unsupervised learning or Self-organization in which an (output) unit is trained to respond to clusters of pattern within the input.

In this paradigm the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli.

Reinforcement Learning This type of learning may be considered as an intermediate form of the above two types of learning. Here the learning machine does some action on the environment and gets a feedback response from the environment. The learning system grades its action good (rewarding) or bad (punishable) based on the environmental response and accordingly adjusts its parameters.

III.   NEURAL NETWORKS IN DATA MINING:

In more practical terms neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. Using neural networks as

a tool, data warehousing firms are harvesting information from datasets in the process known as data mining. The difference between these data warehouses and ordinary databases is that there is actual anipulation and cross-fertilization of the data helping users makes more informed decisions.

Neural networks essentially comprise three pieces: the architecture or model; the learning algorithm; and the activation functions. Neural networks are programmed or "trained" to ". . . store, recognize, and associatively retrieve patterns or database entries; to solve combinatorial optimization problems; to filter noise from measurement data; to control ill-defined problems; in summary, to estimate sampled functions when we do not know the form of the functions." It is precisely these two abilities (pattern recognition and function estimation) which make artificial neural networks (ANN) so prevalent a utility in data mining.

3.1. Feedforward Neural Network:
One of the simplest feed forward neural networks (FFNN), such as in Figure, consists of three layers: an input layer, hidden layer and output layer. In each layer there are one or more processing elements (PEs). PEs is meant to simulate the neurons in the brain and this is why they are often referred to as neurons or nodes.

Input layer   **Hidden layer**   **Output layer**
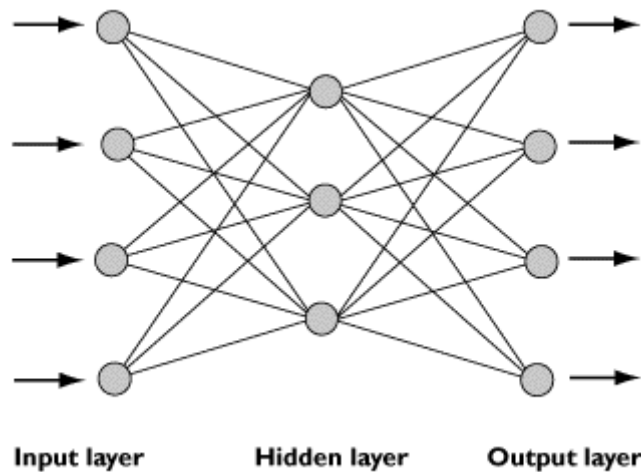
A PE receives inputs from either the outside world or The simplified process for training a FFNN is as follows:
1. Input data is presented to the network and propagated through the network until it reaches the output layer. This forward process produces a predicted output.
2. The predicted output is subtracted from the actual output and an error value for the networks is calculated.
3. The neural network then uses supervised learning, which in most cases is back propagation, to train the network. Back propagation is a learning algorithm for adjusting the weights. It starts with the weights between the output layer PE's and the last hidden layer PE's and works backwards through the network.
4. Once back propagation has finished, the forward process starts again, and this cycle is continued until the error between predicted and actual outputs is minimized.

IV.   DATA MINING PROCESS BASED ON NEURAL NETWORKS
Data mining process can be composed by three main phases: data preparation, data mining, expression and interpretation of the results, data mining process is the reiteration of the three phases. The data mining based on neural network is composed by data preparation, rules extracting and rules assessment three phases, as shown below :

A. Data Preparation:
Data preparation is to define and process the mining data to make it fit specific data mining method. Data preparation is the first important step in the data mining and plays a decisive role in the entire data mining process. It mainly includes the following four processes:
1. Data Clustering:
Data cleansing is to fill the vacancy value of the data, eliminate the noise data and correct the inconsistencies data in the data.
2. Data Option:
Data option is to select the data arrange and row used in this mining.

3. Data Pre-processing:
Data pre-processing is to enhanced process the clean data which has been selected.
4. Data Expression:

Data expression is to transform the data after preprocessing into the form which can be accepted by the data mining algorithm based on neural network. The data mining based on neural network can only handle numerical data, so it is need to transform the sign data into numerical data. The simplest method is to establish a table with one-to-one correspondence between the sign data and the numerical data. The other more complex approach is to adopt appropriate Hash function to generate a unique numerical data according to given string. Although there are many data types in relational database, but they all basically can be simply come down to sign data, discrete numerical data and serial numerical data three logical data types.

B. Data Preparation:

There are many methods to extract rules, in which the most commonly used methods are LRE method, black-box method, the method of extracting fuzzy rules, the method of extracting rules from recursive network, the algorithm of binary input and output rules extracting (BIO-RE), partial rules extracting algorithm (Partial-RE) and full rules extracting algorithm (Full-RE).

C. Rules Assesment:

1. Although the objective of rules assessment depends on each specific application, but, in
    general terms, the rules can be assessed in accordance with the following objectives:
2. Find the optimal sequence of extracting rules,making it obtains the best results in the given
    data set;

3. Test the accuracy of the rules extracted;
4. Detect how much knowledge in the neural network has not been extracted; Detect the
    inconsistency between the extracted rules and the trained neural network.

V.   CONCLUSION

In this paper, we present research on data mining based on neural network. At present, data mining is a new and important area of research, and neural network itself is very suitable for solving the problems of data mining because its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. The combination of data mining method and neural network model can greatly improve the efficiency of data mining methods, and it has been widely used. It also will receive more and more attention.

REFERENCES

[1]   IEEE Transactions on Neural Networks; "Data Mining in a Soft Computing Framework: A Survey", Authors: Sushmita Mitra, Sankar K. Pal and Pabitra Mitra. (January 2002, Vol. 13, No. 1)
[2]   Using Neural Networks for Data Mining: Mark W. Craven, Jude W. Shavlik
[3]   Data Mining Techniques: Arjun K. Pujari
[4]   Introduction to the theory of Neural Computation: John Hertz, Anders Krogh, Richard G. Palmer
[5]   Elements of Artificial Neural Networks: Kishan Mehrotra, Chilukuri K. Mohan, Sanjay Ranka.