

NEURAL NETWORKS IN DATA MINING

S.Sathyamoorthy,

Assistant Professor,

Department of Computer Science/CA/IT,

Bharathiar University Arts and Science College, Gudalur, The Nilgiris, Tamilnadu, India.

Abstract- Data mining is extracts the knowledge/ information from a large amount of data which stores in multiple heterogeneous data base. Knowledge /information are conveying the message through direct or indirect. 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.

Keywords: Data mining, Artificial Neural Network, Feed forward neural networks.

1. 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. The development of Information Technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. Data mining is a powerful concept for data analysis and process of discovery interesting pattern from the huge amount of data, data stored in various databases such as data warehouse, world wide web, external sources.

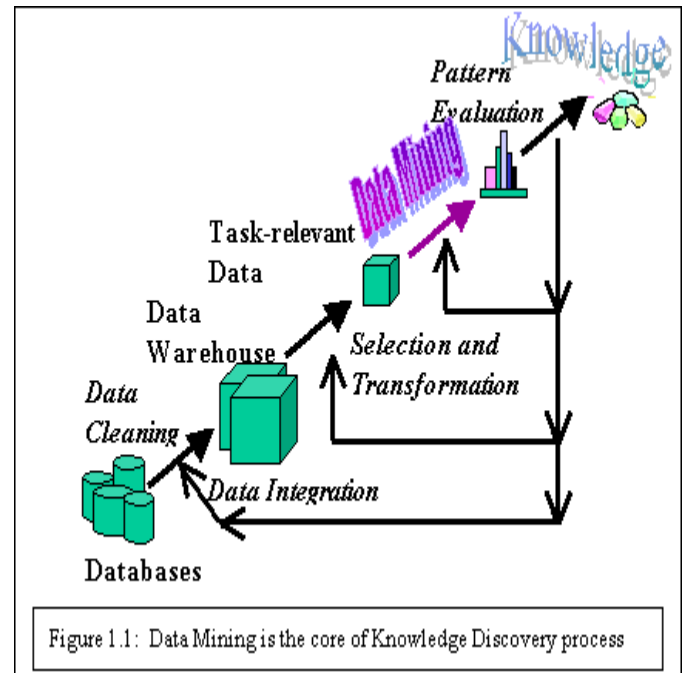
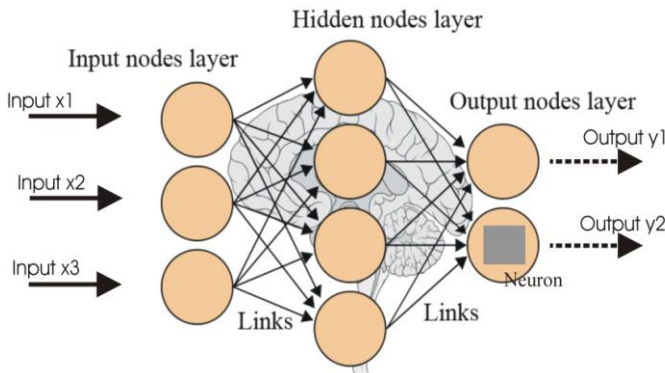


Figure 1.1: Data Mining is the core of Knowledge Discovery process

2. ARTIFICIAL NEURAL NETWORKS:

An artificial intelligence is a very general term but defining it precisely is very difficult. And the design of an artificially intelligent agent totally depends on the fact how we define the term 'Artificial Intelligence'. 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.



Natural neurons receive signals through synapses located on the dendrites or membrane of the neuron. When the signals received are strong enough (surpass a certain threshold), the neuron is activated and emits a signal through the axon. This signal might be sent to another synapse, and might activate other neurons.

2.1 Why use neural networks?

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions. Other advantages include:

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
2. Self-Organization: An ANN can create its own organization or representation of the information it receives during learning time.
3. Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.
4. Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance.

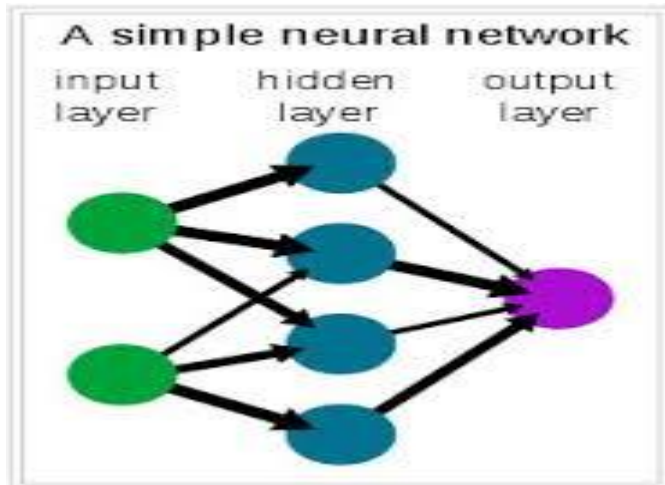
2.2 Neural networks versus conventional computers

Neural networks take a different approach to problem solving than that of conventional computers. Conventional computers use an algorithmic approach i.e. the computer follows a set of instructions in order to solve a problem. Unless the specific steps that the computer needs to follow are known the computer cannot solve the problem. That restricts the problem solving capability of conventional computers to problems that we already understand and know how to solve. But computers would be so much more useful if they could do things that we don't exactly know how to do. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem. They cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable.

Neural networks and conventional algorithmic computers are not in competition but complement each other. There are tasks more suited to an algorithmic approach like arithmetic operations and tasks that are more suited to neural networks. Even more, a large number of tasks, require systems that use a combination of the two approaches (normally a conventional computer is used to supervise the neural network) in order to perform at maximum efficiency.

3. NEURAL NETWORKS IN DATA MINING:

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 Applications of Neural Networks in Data Mining 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 manipulation and cross-fertilization of the data helping users makes more informed decisions.



3.1. Feed forward 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. 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.

4. 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. 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. Rules extracting:

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, partial rules extracting algorithm and full rules extracting algorithm.

C. Rules Assessment:

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.
5. Data mining has importance regarding finding the patterns, forecasting, discovery of knowledge etc., in different business domains. Artificial Neural Networks offer qualitative methods for business and economic systems that traditional quantitative tools in statistics and econometrics cannot quantify due to the complexity in translating the systems into precise mathematical functions. 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.

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] Berson, "Data Warehousing, Data-Mining & OLAP", TMH
- [4] Data Mining Techniques: Arjun K. Pujari
- [4] Introduction to the theory of Neural Computation: John Hertz, Anders Krogh, Richard G. Palmer
- [5] Bradley, I., Introduction to Neural Networks, Multinet Systems Pty Ltd 1997.

[6] Elements of Artificial Neural Networks: Kishan Mehrotra, Chilukuri K. Mohan, Sanjay Ranka.

[7] Industry Application of data mining, <http://www.pearsonhighered.com/samplechapter/0130862711.pdf>

[8] David L Olson, Dursun Delen "Advance data mining techniques" springer 2008

BIOGRAPHY



S.Sathyamoorthy was born in Erode, Tamilnadu, India on 06-05-1981. He received his M.Sc degree in Information Science & Management from Kongu Arts & Science College, Erode, Tamilnadu, India, in 2004. He is now working as an Assistant Professor in the department of Computer science/CA/IT at Bharathiar University Arts and Science College, Gudalur, The Nilgiris, Tamilnadu, India .