

Neural Networks in Data Mining

Ripundeeep Singh Gill, Ashima
Department of Computer Science and Engineering
^{1,2} (Lovely Professional University, Phagwara)

Abstract: - An Artificial Neural Network is an intelligent system which comprises of different components exhibit the human intelligence in the machines. This intelligence comes from the tendency of observation that an artificial neural network have. This much intelligence can be advantageous for the data mining. Artificial neural network is implemented in data mining and its process. After studies, we have found that it has produced very efficient and effective results in the field of data mining. We are presenting a review paper on this appropriate study and in this paper we will discuss about data mining, artificial neural network, its functionality, its relation with biological neural network and role of artificial neural network in the data mining. We will also discuss results of some implementations that has been done in this particular field to justify the effectiveness of artificial neural network in data mining.

Keywords: - ANN (Artificial Neural Network), BNN (Biological Neural Network), Data Mining, KDD (Knowledge Discovery of Data), RNN (Replicator Neural Network).

I. INTRODUCTION

1.1. Data mining

Data Mining is Knowledge Discovery from Data (KDD). We can say it is the part of KDD process. We all have heard the quote “Necessity is the mother of Invention”. This quote is true for Data Mining. An obvious question brighten up into the mind is “Why Data Mining?” The above stated quote is the answer to this question. Today’s world is a technical world due to growth of computer technology and its data. Social Sites, Running Machines, Data Centers, and other technologies produce enormous amount of data. We can say, “We are living in the information age; however we are living in the data age” [1]. Now this much amount of data is difficult to analyze for particular data set. To extract the useful data from huge amount of data, we use KDD process in which Knowledge mining occur. Therefore,

“Extracting useful or knowledgeable data to use from a huge amount of raw data is known as Data Mining. It is also known as Knowledge mining from data”.

KDD process is shown in Fig 1. The whole KDD process is a divided into following steps:

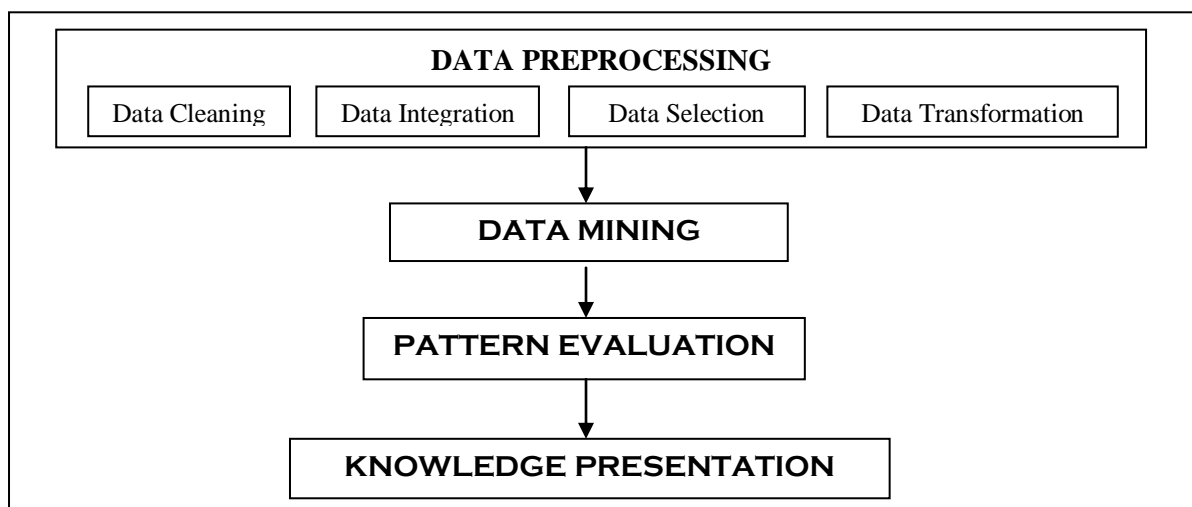


Fig 1: KDD Process

1.1.1. Data Cleaning:

Remove noise and inconsistency from the data

1.1.2. Data Integration:

No. of sources of data are combined together

1.1.3. Data Selection:

Selection of relevant data

1.1.4. Data Transformation:

Modification of data and consolidation of data into form appropriate for mining by performing aggregation operation occur.

1.1.5. Data Mining:

Extraction of useful data from the data set having after data preprocessing i.e. Data Cleaning, Data integration, Data Selection and Data Transformation.

1.1.6. Pattern Evaluation:

Measure of Interestingness, Investigation of pattern that represents knowledge based upon Interestingness Measure. Pattern is interesting if and only if it is valid on test data with some degree of certainty, novel, potentially useful and easily understood by humans [1].

1.1.7. Knowledge Presentation: Present mined data to the user.



Fig 1.2: Data Mining

So here we can observe that data mining is the important step of KDD process; because it gives the expected pattern of data to analyze. Figure 1.2 shows the data mining. Name ‘Data Mining’ referred from ‘Gold Mining’ in which sand is passed through several phases to extract gold from it. The unstructured enormous data is just like sand and useful pattern are analogous to the gold.

1.2. Artificial Neural Network

Artificial neural networks are composed of artificial neurons and interconnection between them. It is based upon the mathematical computation. It is also inspired from the biological neural network consisting in human body, responsible for the response of human body toward potential action applied to it. Artificial Neural Networks, firstly, was invented by Frank Rosenblatt in 1958 at Cornell University. Name of the First Artificial neural networks was “Perceptron”. It is a very effective outcome of the Artificial intelligence to produce effective results. It mimics the human brain except common sense. It is an intelligent system which has capability of the observation. It considered as an incredible network which is modeled as human brain with approximate intelligence of it.

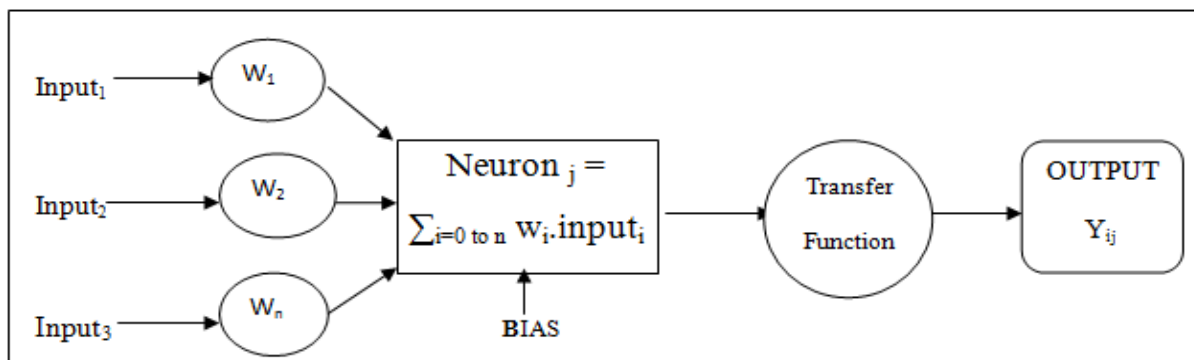


Fig 1.3: Artificial Neural Network

Figure 1.3 represents an Artificial Neural Network and a brief idea that how does it works. In the further section, we will discuss how such system can be used in data mining. Also it will be appropriate to tell that ANN can be used as a decision support system as well as it can be used for image compression, stock market prediction, Character recognition, Traveling Salesman Problems, Medicine, Security, Loan Prediction and many other fields. It shows that neural network is considered as a useful tool and has very broad application

area. In the very next section, we will first discuss how the ANN is related to BNN. We will discuss the functionality of the BNN as well as of ANN. Some of the advantages of Artificial Neural Networks are given below:

- High Accuracy
- Noise tolerance
- Independence from prior assumption
- Ease of maintenance
- Implementation in parallel hardware

II. ARTIFICIAL NEURAL NETWORK AND BIOLOGICAL NEURAL NETWORK

Artificial neural networks are inspired from biological neural networks. Before we see that how neural networks actually work or how it process the given input to give a valuable response; we will study a brief how a biological neural network works in terms to produce response due to potential action upon the body. Our human body consists of billions of neurons and trillion of interconnection between them; hence we can say, a huge neural network exists inside our body. This neural network is the core of our different senses. Let us see how it process the potential action applied to the body.

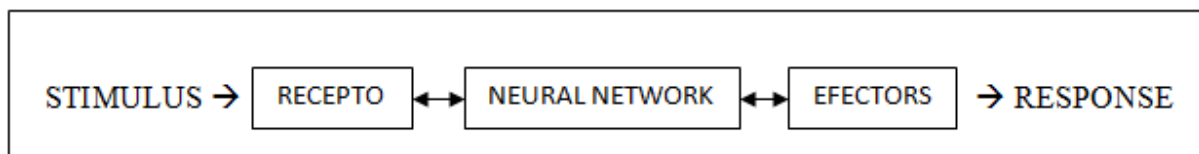


Figure 2.1: Block Diagram of Nervous System

Human brain is complex, non-linear and parallel computer. It is controlled by a nervous system. Figure 2.1 represents the Block Diagram of Representation of Nervous System.

- Stimulus is the potential action applied to the body
- Receptor converts the action into electrical impulse and conveys the information toward neural networks.
- Neural Networks process the information and compute the results and proceed to the effectors.
- Effectors convert the electrical impulse into visible response.
- And then the response is generated.

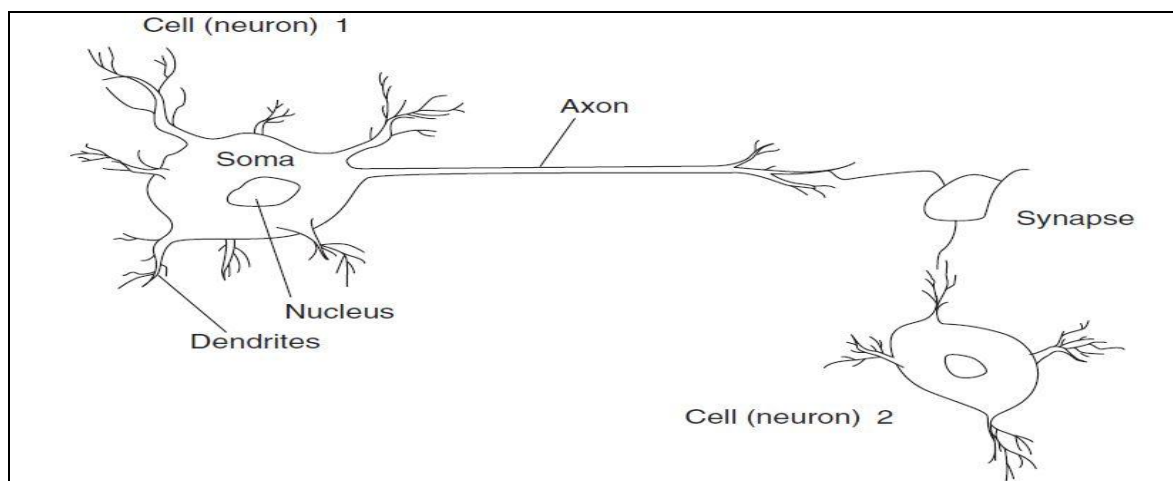


Figure 2.2: Biological Neural Network [3]

In Figure 2.2, there is representation of biological neural networks. If we take a look at biological neural network, the Soma is the neuron node to which Dendrites is the Input and Axon is the output. Synapse acts like synaptic weight for the processing. Axon of one neuron is the dendrites of the other neuron. Synapse also connects the neurons. The term weight is a very important factor for the artificial neural networks. We will discuss about it later. In the Biological neural network, Input is generated by the presynaptic neurons. There is the part of synaptic structure contained within the presynaptic neuron is known as Presynaptic Terminal. Same as the output become input to the postsynaptic neuron, and it contain postsynaptic compartment. Now there is fluid between the gap of the presynaptic terminal and postsynaptic compartment is known as the synaptic cleft. Transmission of the information between presynaptic and postsynaptic neurons is done by chemical signaling.

Presynaptic terminal gives out the neurotransmitter, a kind of molecules, well known as NTXs into the cleft. Receptor protein over the surface of postsynaptic compartment bind the NTXs which cause the electrochemical reaction take place in postsynaptic cell. Biological basis of the weight is as the measure of amount of postsynaptic response that occurs when an input enters the presynaptic terminal. Now with the inspiration of the biological neural networks, the design of artificial neural network came into the mind and a study make this conceptual design a real life practical design which is used in the various fields. Now if we take a look at Figure 1.3, we can find that the functionality is very close to the biological neural networks. The motive behind the design of neural networks is to provide the human intelligence in the computer science to get more effective results. As we have discussed the advantages of artificial neural networks, it is explicit to say that it can produce effective results. The basic components and requirements of the Artificial Neural Networks are:

- Weight
- Artificial Neurons
- Transfer Function

These three components or requirements are the bases of the Artificial Neural networks. In brief steps, I would like to mention the functionality of the Artificial Networks:

- Input is given to the Artificial Neuron.
- Observing the Input, it generates random weight over the links.
- In the node, a summation is performed over the multiplication of generated weight and provided input.
- Some bias over the node also influence the result generated by node due to surroundings.
- Then the transfer function is applied to the generated output of node and provides the output.
- This output can be the input to the further joined network of neurons.

Term “Weight” is the important factor, as I said earlier, because the whole output is dependent upon this factor. When input is introduced to the neuron, this weight generated by observing the input act like synapse in biological neural network. Therefore it is also known as the synaptic weight of neural network. Now synaptic weight, when first output gets generated, will be adjusted within accordance of expected output. The adjustment of weight is done by different algorithms. A very popular algorithm to adjust the weight is back-propagation algorithm. After feeding more and more training data, when the output become accurate as expected, Adjustment then stopped, otherwise adjustment of weight continuously run on until it give expected results. Tables are given below to show analogy of the biological neural network and artificial neural network. We can observe after a brief study that Artificial Neural Network has brought the human intelligence into the computers except common sense. Human Intelligence is quite enough to take good decisions; therefore, we can say Artificial Neural Network can be fruitful decision support system. According to our point of view, data mining process is a kind of decision making about data or information, because during KDD process, we need to decide what kind of data is useful and what kind of data is futile. For example, if we are gathering the detail of a person then it is highly depend upon the reason behind it. Because it will make us able to decide that what kind of information we need to collect to fulfill that specific reason. In the very next section, we will study about how Artificial Neural Networks are beneficial in the field of data mining.

Biological Neural Network	Artificial Neural network
Soma	Node
Dendrites	Input
Axons	Output
Synapses	Weight
Many Neurons	Few Neurons

TABLE 2.1: Nomenclature BNN and ANN components and its relation [3]

Parameters	Biological Neural Network	Artificial Neural Network
Number of neurons	10^{10}	$\wedge 000$
Number of Synapses	60×10^{12}	$\wedge 000$
Speed[s/op]	10^{-3}	10^{-9}
Energy[J/op]	10^{-16}	10^{-6}

TABLE 2.2: Parameters of biological and artificial neural network [5]

Table 2.1 represents the nomenclature adopted in artificial neural network on the basis of the biological neural network. This nomenclature symbolizes the relationship between the biological neural network as well as similarity between component and action exists in both the architecture. Table 2.2 represents the comparison between biological neural network and artificial neural network on the basis of some parameters.

III. ARTIFICIAL NEURAL NETWORKS IN DATA MINING

Artificial neural network is considered as a tool which is highly beneficial for data mining. Artificial Neural Network is also known as non linear statistical data modeling tools which recognize the appropriate pattern and respond with an appropriate output. Artificial neural networks are good at observation due to the synaptic weight generated by it; so high accuracy and performance can be expected. Artificial Neural Network consists of three parts:

- Architecture : how neurons are organize and how they interact with each other
- Learning Algorithm : An algorithm that train the ANN to perform specific task
- Activation Function: An activation function or also known as transfer function that is performed over the output generated by node.

Artificial Neural networks are capable to Store data, Recognize pattern and Retrieve the fruitful pattern. Today's age is the information age due to massive growth of data. Massive growth of data is due to growth of technology. We know that a running jet engine can produce gigabyte of data within few minutes. From this we can trace an idea that how much data is produced by technologies used on the daily basis. So it is very necessary to find out the value-able data to operate. It is bewilder to do this job manually so automated processing is needed which can be effectively done by an intelligent system; so artificial neural network is.

Data mining is either descriptive data mining in we observe what is happening inside the data without having any kind of idea or predictive data mining in which we observe the patterns based upon the history of data. In both cases artificial neural network is the appropriate tool for the performance. Data mining basically consists of different phases:

- Pattern Recognition
- Classification
- Clustering
- Outlier detection

Pattern recognition is the task which is either dependent upon what is happening or what is the history of data. Artificial neural network is considered good at observation; so pattern recognition can be well performed by the artificial neural network. Artificial neural network has been used pattern recognition and it has produced admirable results for example in voice pattern recognition, it recognize with rate of 70%, for noisy data, up to the 99%.[7] Artificial neural network has also given appropriate results in the association and correlation intend to find frequent pattern.

Classification is the main process of data mining in which we classify the data according to the different classes, for example, Genre classification of musical data is the necessity of today's world and Artificial neural network is used for this purposes and Implementation of Neural network has achieved 85% of the accuracy in musical data classification by genre; also, 65% of the accuracy in classification of sub genre. [9]

Clustering the data means we make clusters of similar data after classify it. Artificial neural network is also beneficial for the purpose. ART1 (Adaptive Resonance Theory 1) neural network approach for clustering the data has given the advantages of Flexibility, high speed computation, and handling large input vector. [10]

Outlier detection is the method to detect an outlier which does not belong to a specific bunch. Artificial Neural Networks are also good at outlier detection. Let us consider an example of replicator neural network for the outlier detection. RNN provide the 80.1% rate in network intrusion detection. [11]

So after these studies we can say that artificial neural network is a useful tool for data mining. We have considered one example each but there is a lot more to see. These researches have proved ANN as an appropriate tool for mining of data. ANN has also been used in financial fraud detection, Apriori algorithm, classification in medical field, loan prediction, financial forecasting, stock market, weather and rainfall forecasting and many more. Interesting fact is that ANN has produced effective results in each implementation.

IV. CONCLUSION

Artificial neural network is an intelligent system that is used in data mining for different processes like pattern recognition, classification, frequent item set, clustering, outlier detection etc., and it has produced effective and efficient results which are enough to prove the artificial neural network as a useful tool in data mining. We can observe that how an intelligent system can produce effective results in a particular field. Data mining is so important now days. As we have discussed in paper that technology is growing rapidly and producing a large amount of data in milliseconds, so to process that data we need some rapid system. There we can see the importance of neural network to do specific job and help us to perform the rest of the task. We have

concluded that it is a helpful tool that can be use for the purpose of data mining as well as we can also implement it to the new field of several industries. As data mining is important to several fields, so the neural network also acquire the same importance. There is always a scope of improvement in everything, therefore improvement or enhancement can be done to get better or best results.

V. ACKNOWLEDGEMENT

Review paper is supported and guided by Mr. Hem Kumar, Assistant Professor at Lovely Professional University. We are highly obliged to their support and guidance. Special thanks to Mr. Sashi Kanth Rathore, Assistant Professor at Lovely Professional University and Mrs. Jaspreet Kaur, Assistant Professor at Lovely Professional University for their valuable support and encouragement.

REFERENCES

- [1] M. Kamber, J. Pei, J. Han, *Data Mining: Concepts and Techniques*, (Morgan Kaufmann, p. 560.)
- [2] S. Haykins, *Neural Network - A comprehensive foundaion (2nd Edition)*, (Prentice Hall, 1999).
- [3] J. E.Aronson, T.-P. Leing, E. Turban, "Neural Network in Data Mining," in *Decission Support System*, (Prentice-Hall of India Private Limited, 2007).
- [4] R.Wells, "Synaptic Weight Modulation and Adaptation," May 15,2003.
- [5] M.Hajek, "Neural Networks," 2005.
- [6] A. S. Chauhan, Dr. Y. Singh, "Neural Network in Data Mining," *Journal of Theoretical and Applied Information Technology*, vol. 5, p. 6, 2005-2009.
- [7] D. Bhattacharyya, T. H. Kim. J. Kumar, "Use of Artificial Neural Network in Pattern Recognition," *Internationall Journal of Software Engineering and its Application*, vol. 4, 2010.
- [8] P. Scott, "Music Classification using Neural Network," *Spring*, 2001.
- [9] C. Mckay, "Using Neural Network for Musical Genre Classification".
- [10] R. Joshi, N.Kumar, "Data Clustering Using Artificial Neural Networks," in *Challenges and Opportunities in Information Technology*, 2007.
- [11] H. He, G. William, R. Baxter, Simon Hawkins, "Outlier Detection Using Replicator Neural Networks".