IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 5 Issue 4, April 2018
ISSN (Online) 2348 – 7968
www.ijiset.com

Artificial Neural Network Based Data Mining

Fathima Safna P.A [1], Rekha Sunny T[2]

[1] Student, MCA Department, SCMS, Cochin, Kerala, India

[2] Assistant Professor, MCA Department, SCMS, Cochin, Kerala, India

Abstract

Explosion of Big data resulted in the huge requirement for efficient data mining techniques that can work upon large data that is varied and extensive in its very nature and content. Neural networks play a major role in developing these algorithms due to its robustness, parallelism, high affordability to the noise data and low error rate. This paper gives an overview of artificial neural network, how neural networks can be used to improve the efficiency of data mining techniques and the different neural network-based data mining types.

Keywords: *Data Mining, Neural Networks, Data Mining Process, Knowledge Discovery*

1. Introduction

Data mining is the process of analyzing hidden patterns of data according to different perspectives, for categorization into useful information. The data is collected and assembled in common areas, such as data warehouses, for efficient analysis. Data mining tools predict future trends and behaviors, thus allowing businesses to make proactive, knowledge -driven decisions. Data mining principles have been around for many years, but, with the advent of big data, it is even more prevalent. It caused an explosion in the use of more extensive data mining techniques, partially because the size of the information is much larger and because the information tends to be more varied and extensive in its very nature and content [11].

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. In more practical terms neural networks are non-linear statistical data modelling tools [1].

In data mining neural network methodology is used for classification, clustering, feature mining, prediction and pattern recognition. It imitates the neurons structure of animals. Through training data mining, the neural network method gradually calculates the weights the neural network connected. The neural network model can be broadly divided into three types such as feed forward networks, back propagation algorithm and Self-organization networks. Feed forward networks regards the perception backpropagation model and the function network as representatives, and mainly used in the areas such as prediction and pattern recognition. Backpropagation, or propagation of error, is a common method of teaching artificial neural networks how to perform a given task. The back-propagation algorithm is used in layered feedforward ANNs. Self-organization networks regard adaptive resonance theory (ART) model and Kohonen model as representatives, and mainly used for cluster analysis [7].

The rest of this paper is organized as follows: Section 2 provides an overview of data mining and artificial neural network techniques. Section 3 discusses artificial neural network-based data mining. In section4 different data mining types based on neural networks are discussed. In section 5, the comparison of self-organization neural network and fuzzy neural network is done and section 6 concludes the paper.

2. Literature Review

2.1 Data Mining

Data mining is the process of extraction of hidden predictive information from large databases. It is a powerful technology and has great potential to help companies focus on the most important information in their data warehouses. Databases are exploited for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

2.1.1 Data Mining Process

The data mining process is often characterized as a multi-stage iterative process involving data selection, data cleaning, application of data mining algorithms,

evaluation, and so forth. Here we adopt a somewhat different process-oriented view and break it down into five basic steps.

1.Exploring and Preprocessing: This consist of initial steps in exploring, visualizing, and querying the data, gaining insight into the data in an interactive manner. It can also include variable selection, data focusing, and data validation can also be included in this initial step. Data preprocessing transforms the data into a format that will be more easily and effectively processed for the purpose of the user.

2.Modelling: This consist of the steps involved in (a) selecting the model representations that we seek to fit in to the data (e.g., a tree, a linear function, a probability density model, etc.), (b) selecting the score functions that score different models with respect to the data, and (c) specifying the computational methods and algorithms to optimize the score function (e.g., greedy local search). These “components" combined together specify the data mining algorithm to be used. The components may be “precompiled" into a specific algorithm.

3.Mining: This is the step (often repeated) of actually running a particular data mining algorithm on a particular data set.

4.Evaluating: This step (often ignored) of critically evaluating the quality of the output of the data mining algorithm from step 3, both the predictions of the model and the interpretation of the fitted model itself.

5.Deploying: This step (rarely achieved) of putting a model from a data mining algorithm into routine predictive use, e.g., using the model continuously in real-time for scoring customers visiting an ecommerce Web site. A challenging (and under-appreciated) technical issue in this context is how and when models should be updated for such continuous data stream applications .

2.1.2 Data Mining Techniques

There are several techniques used in data mining that describe the type of mining and data recovery operation. Some of the key techniques are as follows:

Association (Rule Mining): Association is probably the better known and straightforward data mining technique. A simple correlation between two or more items is made, mostly of the same type in order to identify patterns. For example, while tracking customers buying habit, you might identify that a customer always buys chocolate when they buy cream, hence we can suggest that next time when they buy cream, they might also need chocolate.

Classification: Classification can be used to build up an idea of the type of customer, item, or object by describing multiple attributes to identify a particular class. For example, we can easily classify cars into different types (sedan, 4x4, convertible) by identifying different attributes (number of seats, car shape, driven wheels). Given a new car, you might apply it into a particular class by comparing the attributes with our known definition. You can apply the same principles to customers, for example by classifying them by age and social group. Additionally, we can use classification as a feeder to, or the result of, other techniques.

Clustering: By examining one or more attributes or classes, we can group individual pieces of data together to form a structure opinion. At a simple level, clustering is using one or more attributes as your basis for identifying a cluster of correlating results [11].

2.2 Artificial Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. Artificial Neural Networks, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well.

In human body work is done with the help of neural network. Neural Network is just a web of inter connected neurons. With the help of this interconnected neurons all the parallel processing is done in human body. It is composed of a cell body or soma and two types of out reaching tree like branches: the axon and the dendrites. The cell body has a nucleus that contains information about hereditary traits and plasma that holds the molecular equipment or producing material needed by the neurons. The whole process of receiving and sending signals is done in particular manner like a neuron receive signals from other neuron through dendrites. The Neuron send signals at spikes of electrical activity through a long thin stand known as an axon and an axon splits this signals through

synapse and send it to the other neurons as shown in figure 1. [9]

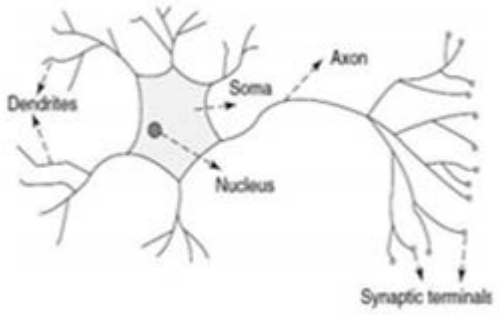

Figure 1: Human Neuron

An artificial neuron is a device with many inputs and one output as shown in figure 2 [9].

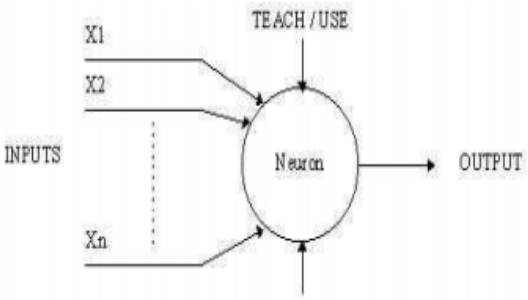

Figure 2: Artificial Neuron

Similar to biological neuron artificial neural network contain artificial neurons and they can also receive inputs from other elements or other artificial neurons. Once the inputs are weighted and added, the result is then transformed by a transfer function into an output [9].

2.2.1 Training Of Artificial Neural Network

A neural network can be configured such that by providing a set of inputs it provides the desired output. Various methods exist to set the strength of connections. 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. The learning situations are categorized into 3.

Supervised learning: In this method the network is trained by giving input and matching output patterns. The inputs are then processed and the resulting output is then compared with the desired output pattern. Errors are then propagated back through the system, causing the system to adjust the weights which control the network. It is also known as associative learning.

Unsupervised learning: In this method an output unit is trained to respond to a 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.It is also known as adaptive learning or self oraganization.

Reinforcement learning: In this kind of learning, 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 parameter as an intermediate form of the supervised and unsupervised learning [9].

2.2.2 Neural Network Techniques

Feedforward Neural Network: The simplest feedforward network consist of three layers as shown in figure 3 : an input layer, a hidden layer and an output layer. Each layer in this network consist of one or more processing elements. Processing elements is meant to simulate the neurons in the brain and this is why they are often referred to as neurons or nodes. A processing element receives inputs from either the outside world or the previous layer. There are connections between the PEs in each layer that have a weight associated with them. This weight is adjusted during training. Information only travels in the forward direction through the network - there are no feedback loops as shown in figure3 [9] .

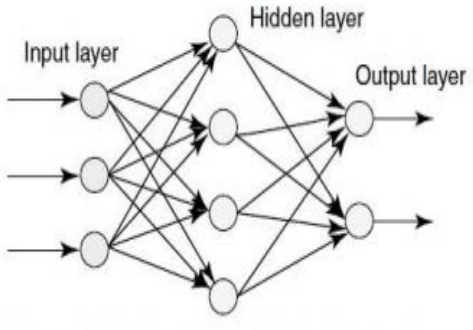

Figure 3: Layered Neural Network

The Back Propagation Algorithm: Backpropagation, or propagation of error, is a common method of teaching artificial neural networks how to perform a given task. The back propagation algorithm is used in layered feedforward ANNs. This means that the artificial neurons are organized in layers, and send their signals forward, and then the errors are propagated backwards. The back propagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then the error (difference between actual and expected results) is calculated. The idea of the back propagation algorithm is to reduce this error, until the ANN learns the training data [3].

3. Artificial Neural Network Based Data Mining

There are three main phases in data mining and they are: data preparation, data mining, expression and interpretation of the results. Data mining process is the reiteration of the three phases. The details are as shown in figure 4[3]:

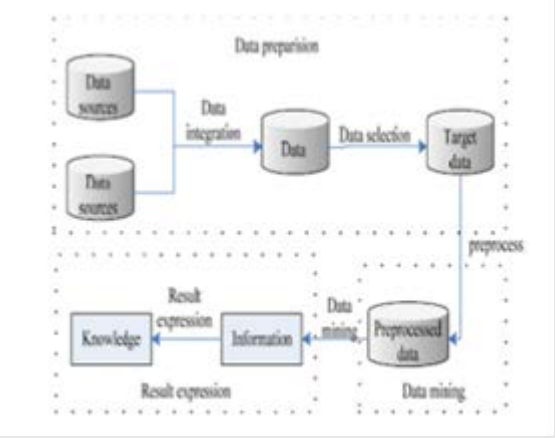

Figure 4: Data mining process

The data mining based on neural network consist of three phases and they are: data preparation, rules extracting and rules assessment. Figure of data mining process based on neural network is as follows as shown in figure 5[3].

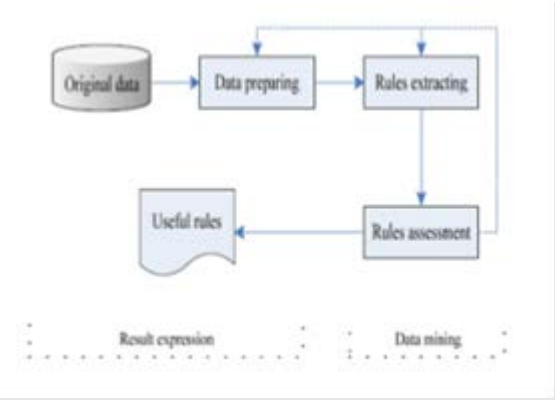

Figure 5: Data mining process based on neural network

3.1. Data Preparation

Data preparation is done to define and process the mining data to make it fit specific data mining method. The first important step in the data mining is data processing and it plays a decisive role in the entire data mining process. This consist of the following four processes.

1) Data cleaning: Data cleansing is done to fill the vacancy value of the data, eliminate the noise data and to correct the inconsistencies in the data.

2) Data option: Data option is done to select the data arrange and row used in this mining [3].

3) Data preprocessing: This technique is used to process the raw data into an understandable format. It transforms the data into a format that can be used for efficient mining and produce accurate results.

4) Data expression: This is to transform the data after preprocessing into an acceptable form by the data mining algorithm based on neural network. The data mining based on neural network can only handle numerical data, so the sign data needs to be transformed into numerical data. This can be established by creating a table with one-to-one correspondence between the sign data and the numerical data [10].

3.2. Rules Extracting

There are a number of methods to extract rules, and the common methods used 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) [10].

3.3. Rules Assessment

The objective of rules assessment varies with each specific application but, in general terms, the following objectives can be used to assess rules.

1) Find the optimal sequence of extracting rules, making it obtains the best results in the given data set;
2) Test the accuracy of the rules extracted;
3) Detect how much knowledge in the neural network has not been extracted;
4) Detect the inconsistency between the extracted rules and the trained neural network [3].

4. Data Mining Types Based on Neural Network

The data mining types based on neural networks are many, but only two types are commonly used and they are based on the self-organization neural network and the fuzzy neural network.

4.1. Data Mining Based on Self-Organization Neural Network

In this process learning is done without teachers. Through the study, the important characteristics or some inherent knowledge in a group of data, such as the characteristics of the distribution or clustering according to certain feature. Scholars T. Kohonen of Finland considers that the neighbouring modules in the neural network are similar to the brain neurons and play different rules, through interaction they can be adaptively developed to be special detector to detect different signal. Because the brain neurons in different brain space parts play different rules, so they are sensitive to different input modes. T. Kohonen also proposed a kind of learning mode which makes the input signal be mapped to the low-dimensional space, and maintain that the input signal with same characteristics can be corresponding to regional region in space, which is the so-called self-organization feature map [3].

4.2. Data Mining Based on Fuzzy Neural Network

The fuzzy neural networks frequently used in data mining are fuzzy perception model, fuzzy BP network, fuzzy clustering Kohonen network, fuzzy inference network and fuzzy ART model. Here the fuzzy BP network is developed from the traditional BP network. In the traditional BP network, if the samples belonged to the first k category, then except the output value of the first k output node is 1, the output value of other output nodes all is 0, that is, the output value of the traditional BP network only can be 0 or 1, is not ambiguous. However, in fuzzy BP networks, the expected output value of the samples is replaced by the expected membership of the samples corresponding to various types. After training the samples and their expected membership corresponding to various types in learning stage fuzzy BP network will have the ability to reflect the affiliation relation between the input and output in training set and can give the membership of the recognition pattern in data mining. Fuzzy clustering Kohonen networks achieved fuzzy not only in output expression, but also introduced the sample membership into the amendment rules of the weight coefficient, which makes the amendment rules of the weight coefficient has also realized the fuzzy. [3]

5. Comparison of Self-Organization Neural Network and Fuzzy Neural Network

Learning algorithm used by self organizing network is called Kohonen algorithm. The algorithm is a non-supervised clustering method, which clusters data by repeatedly learning. Its clustering process also carries on through the certain unit competition current object. Weight vector is closest the unit of current object which becomes active unit or winning unit. In order to close the input object, it is necessary for the winning unit and its nearest neighbor's weight to be adjusted. SOM assumes that the input object has some topological structure or order, the unit organization in unit forms a characteristic mapping [7].
The neural network has strong functions of learning, classification, association and memory. But when we use neural network for data mining, the greatest difficulty is that the output results cannot be intuitively represented. After the introduction of the fuzzy processing function into the neural network, other than increasing its output expression capacity, it can also make the system more stable [3].

6. Conclusion

Data mining is an important area of research, and neural network itself is very suitable for solving the problems of data mining because of its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. Hence in this paper we discuss how efficiency of data mining methods can be improved by combining neural network model with data mining techniques.

References

[1] Nashaat El-Khamisy Mohamed, Ahmed Shawky Morsi El-Bhrawy, "Artificial Neural Networks in Data Mining", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-06,p-ISSN: 2278-8727, Volume 18, Issue 6, Ver. III (Nov.-Dec. 2016).

[2] Priyanka Gaur, "Neural Networks in Data Mining" , International Journal of Electronics and Computer Science Engineering ,www.ijecse.org.
[3] Sujata S.Kharat , Vamshi Krishna , "To Study Artificial Neural Networks in Data Mining and Its Method", Volume 3, Issue 7, July 2015 International Journal of Advance Research.
[4] Sujith Jayaprakash, Ghana Balamurugan E., "A Comprehensive Survey on Data Preprocessing Methods in Web Usage" ,(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015.
[5] Sanjesh Ghore, "Data Mining used of Neural Networks Approach" ,IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6.
[6] Amit Bhagat, "New Effective Data Mining Method Based on Neural Networks" International Journal of Engineering and Technical Research (IJETR) ISSN: 2321-0869, Volume-3, Issue-6, June 2015 .
[7] Guoquan Jianga, Cuijun Zhaob, "The Research of Data Mining Based on Neural Networks" , 2011 International Conference on Computer Science and Information Technology (ICCSIT 2011) IPCSIT vol. 51 (2012), IACSIT Press, Singapore.
[8] David Hand, "Principles of Data Mining" , Massachusetts Institute of Technology, 2001.
[9] Nashaat El-Khamisy Mohamed, Ahmed Shawky Morsi El-Bhrawy, "Artificial Neural Networks in Data Mining" , IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-066, p-ISSN: 2278-8727, Volume 18, Issue 6, Ver. III (Nov.-Dec. 2016), PP 55-59.
[10] Xianjun Ni, "Research of Data Mining Based on Neural Networks", World Academy of Science, Engineering and Technology 39 2008.
[11] Martin Brown, "Data mining techniques", IBM developer works December 11, 2012.