

Research of data mining based on neural networks

ANKITA AGARWAL

CS/IT Department, Sunderdeep College of Engineering and Technology, UPTU

Ghaziabad, Uttar Pradesh, India

Abstract

Data mining is a powerful new technology with great potential to help companies focus on the most important information in the data they have collected about the behavior of their customers and potential customers. It discovers information within the data that queries and reports can't effectively reveal. Data mining is the extraction of implicit, previously unknown, and potentially useful information from data. This paper include various data mining methods like classification , clustering, association etc and Data mining process based on neural network It consists of Backpropogation algorithm and various technique for improving Artificial Neural Network.

Keywords: Data Mining, Neural network, Back propagation algorithm, Neuro fuzzy system

1. Introduction

Data mining refers to process or method that extracts or mines interesting knowledge or patterns form large amounts of data. It searches for relationship and patterns that exist in large database but are hidden among vast amount of data. It is referred as knowledge extraction and data dredging.

Data mining technique are:

1. **Classification:** It is organization of data in given class. It uses a training set where all the objects are already associated with known class label .It is also known as supervised classification.
2. **Clustering:** It refers to grouping of records. Observations into classes of similar objects. It is referred as unsupervised classification. It is based on principle of maximizing similarity between objects in same class and minimizing similarity between objects of different class.
3. **Association:** It studies frequency of items occurring together in transactional database. Support identifies frequent item set. Confidence is condition probability that an item appear in transaction when another item appears.

2. Data mining method

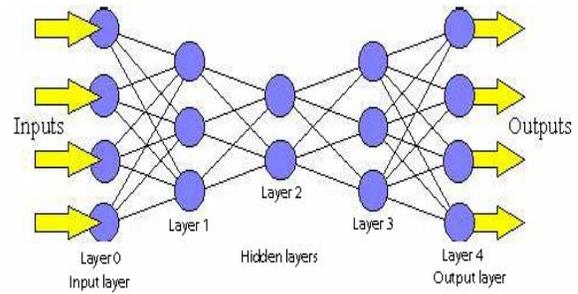
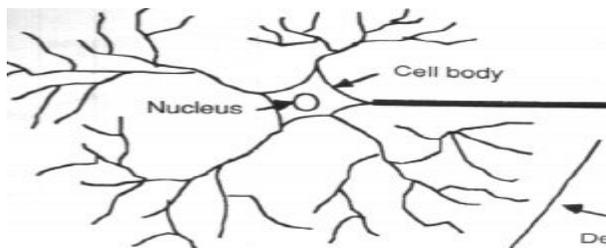
In data mining various methods are used for gaining knowledge.

1. **Decision tree:** It is hierarchical group of relationship organized into tree like

structure. It determines which data and in which order should be collected to achieve effective decision that represent knowledge with minimal cost.

2. **Genetic algorithm:** They are class of randomized search procedure capable of adaptive search over a wide range of search space topology. They have ability to solve problem in parallel, so it is a powerful tool for data mining. Genetic algorithm search uses a set of solutions during each generation.

3. **Neural network :**



It is directed graph consisting of nodes with interconnecting synaptic and activation link. It is a parallel processing network which is generated by stimulating image intuitive thinking of human on basis of research of biological neural network. It is also referred as 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.

3. Neural network method in data mining

In neural network, knowledge representation of surrounding environment is defined by value taken on by parameters such as synaptic weight and biases of network.

Three neural network method.

- a) **Feed forward network:** It regards perception Back propagation model and function network as representative.

- b) **Feed back network :** It regards perception Hopfield discrete model and continuous

Model as representative.

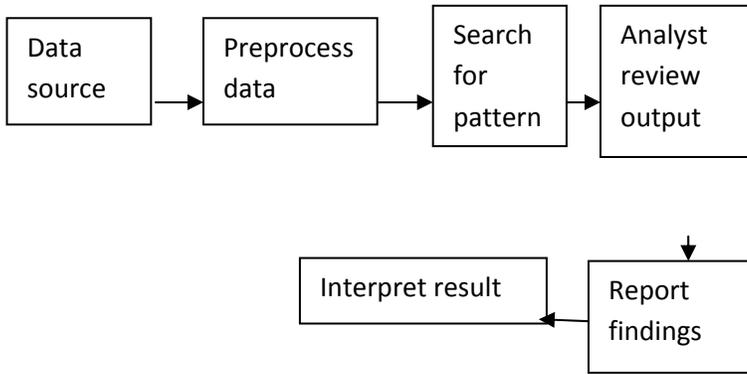
- c) **Self organization network :** It regards Adaptive Resonance theory

and Kohonen model as representative

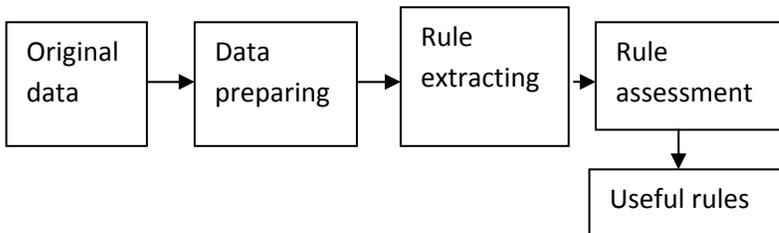
Pattern recognition and function estimation abilities of neural network are used in data mining. Neural network was trained to store, recognize and retrieve pattern or database entries, to filter noise from measurement data.

4. Data mining process based on neural network

Data mining process consist of:



Data mining process based on neural network consist of:



1. Data preparing :

It defines and process mining data to make it fit specific data mining method. It includes:

- a) **Data cleaning:** It routine work to clean data by filling in missing values, smoothing noise data, removing outliers and resolve inconsistency.
- b) **Data option :** It is used to select data arrange
- c) **Data preprocessing:** It describes any type of processing performed on raw data to prepare it for another processing procedure.
- d) **Data transformation:** In this selected data is transformed into a form appropriate for mining procedure. Data mining Based on neural network can handle numeric data only.

2. Rules extracting:

There are various methods for extracting rules. These are mainly

- a) LRE method
- b) Black Box method
- c) BIO-RE

3. Rule assessment :

It is basically done for finding optimal sequence of extracting rules, test accuracy of rules extracted and to detect how much

knowledge in neural network had not been extracted.

5. Neural network learning algorithm

Back propagation is a common method of training artificial neural networks so as to minimize the objective function. It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks. The term is an abbreviation for "backward propagation of errors". Back propagation requires that the activation function used by the artificial neurons be differentiable.

The back propagation learning algorithm can be divided into two phases: propagation and weight update.

Phase1: Propagation

Each propagation involves the following steps:

1. Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations.
2. Backward propagation of the propagation's output activations through the neural network using the training pattern's target in order to generate the deltas of all output and

hidden neurons. Backward propagation of the propagation's output activations through the neural network using the training pattern's target in order to generate the deltas of all output and hidden neurons.

PHASE 2: WEIGHT UPDATE

For each weight-synapse follow the following steps:

1. Multiply its output delta and input activation to get the gradient of the weight.
2. Bring the weight in the opposite direction of the gradient by subtracting a ratio of it from the weight.

This ratio influences the speed and quality of learning; it is called the *learning rate*. The sign of the gradient of a weight indicates where the error is increasing; this is why the weight must be updated in the opposite direction. Back propagation is an iterative process that can often take a great deal of time to complete.

6. Improve Artificial neural network performance

There are basically 2 techniques for improving ANN performance:

a) Using Genetic Algorithm with Neural Network

Neural networks and genetic algorithms are two techniques for optimization and learning, each with its own strengths and weaknesses. Genetic algorithm can be applied to train artificial neural network for classification purpose.

In NNUGA, we are using GA to find the solution to a **classification problem** with a **neural network**. The neural network is a structure which is able to respond with *true* (1) or *False* (0) to a given input vector. We are trying to "teach" our neural network to correctly classify a set of input vectors, which can be thought of as **learning a concept**. We then expect that when the neural network will be presented with a vector P not from this set, it will tend to exhibit **generalization** by responding with an output similar to target vectors for input vectors close to the previously unseen input vector

a) Fuzzy into neural

Fuzzy system involve control written membership function. Rule is

IF(A is small)AND(B is large)then(X is medium)

A and B parts input to system. AND is fuzzy operation. X is operational value for target system Small is LHS membership function and Medium is RHS membership function .

The value of A,B are entered into system , then fuzzyt operation AND of “A is small” and “B is a large” is calculated The result is called the grade for RHS membership function and is then used to multiply medium .All function are mapped and are mapped into neural network.

b) Neuron fuzzy system

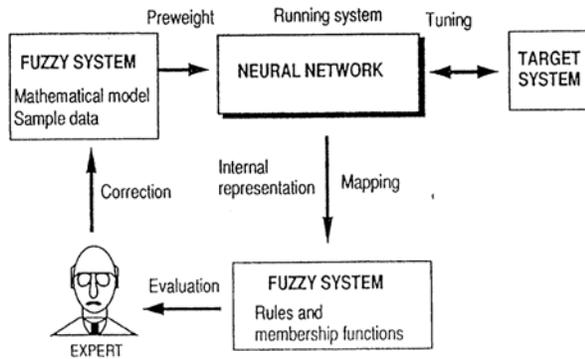


Figure 1. Neurofuzzy system

Neuron fuzzy system proposed enables us to switch us from fuzzy system to neural network and vice versa. This system is able to process and learn numerical information as well as linguistic information.

b)Neural to fuzzy

In this neural network is treated as balck box and we go inside the neural network

Network is assumed to have learned data.Then neural network is mapped into fuzzy system into fuzzy rules.

7. Conclusion

This paper presents research on data mining based on neural network. Neural network solves the problem of data mining as it has parallel processing, distributed storage, high degree of fault tolerance, good robustness. In this paper's Data mining process based on neural network, backpropagation algorithm are explained and emphasis is laid on various techniques for improving the performance like Using Genetic Algorithm with Neural Network and neuro fuzzy system.

References

1. Xianjun Ni, Research of data mining based on neural networks, 2008.
2. Yashpal Singh, Alok Singh Chauhan, Neural Network in data mining, 2005-2009.
3. Haykin, Neural Networks, Prentice Hall International Inc, 1991
4. Jiawei Han and Micheline Kamber, Data Mining : concepts and techniques, 2nd ED
5. K. Amarendra, KV Lakshmi and KV Ramani, Research of data mining using neural networks.