ISSN:0975-9646

Improving Efficiency of Data Mining through Neural Networks

D. Prashanth Kumar[1] , B. Yakhoob[2] , N. Raghu[3]

Assistant Professors in CSE Department
Kamala Institute Of Technology & Science, Singapur

Abstract: **"Data Rich and Information Poor" is the tagline on which the field Data Mining is based on. Data Mining mines the required Knowledge for the vast amount of data that is available form various sources in almost all the areas now a days. From this huge amount of data, required knowledge is to be extracted in the required format. Data Mining however deals with this problem very well. To enhance the capability of Data Mining, we can take help of Neural Networks which will give the accurate and efficient results in some cases.**

Neural network is a parallel processing network which generated with simulating the image intuitive thinking of human, on the basis of the research of biological neural network, according to the features of biological neurons and neural network and by simplifying, summarizing and refining. It uses the idea of non-linear mapping, the method of parallel processing and the structure of the neural network itself to express the associated knowledge of input and output. Initially, the application of the neural network in data mining was not optimistic, and the main reasons are that the neural network has the defects of complex structure, poor interpretability and long training time. But its advantages such as high affordability to the noise data and low error rate, the continuously advancing and optimization of various network training algorithms, especially the continuously advancing and improvement of various network pruning algorithms and rules extracting algorithm, make the application of the neural network in the data mining increasingly favored by the overwhelming majority of users.

General Terms:
Neural networks, Data Mining

Keywords:
Activation Function, Artificial Neural Network, Node

1. INTRODUCTION

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Artificial neural networks are inspired by the operation of the human brain. It is a model of the biological neuron as a circuit component to perform computational tasks. Artificial neural networks consist of a number of simple computing elements called neurons that are modeled after the human nerve cell. Each neuron receives a number of input signals and performs a simple operation on this set of inputs. The output of each neuron is fanned out to the inputs of other neurons.

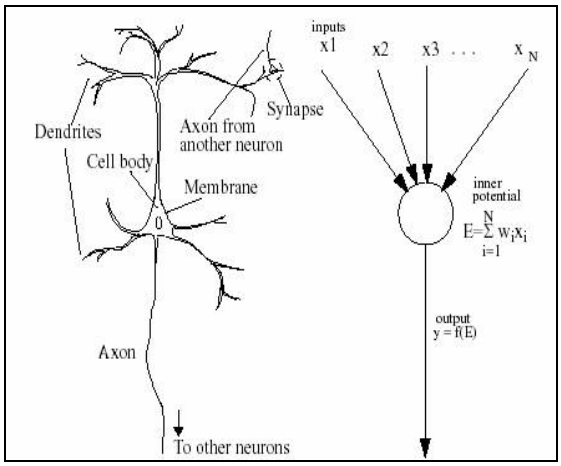


Figure 1 Human nerve system and its artificial equivalent

In Figure 1, a human nerve cell, or neuron, (a) and its artificial equivalent (b) are sketched. The neuron receives a set of input signals via a number of tentacles or dendrites. At the tip of each dendrite the input signal is weighted with a factor w, which can be positive or negative. All the signals from the dendrites are added in the cell body to contribute to a weighted sum of inputs of the neuron. If a weight is positive the corresponding input will have an excitatory influence on the weighted sum. With a negative weight, an input decreases the weighted sum and is inhibitory. In the cell body the weighted sum of inputs is compared to a threshold value. If the weighted sum is above this threshold, the neuron sends a signal via its output to all connected neurons. The threshold operation is essentially a nonlinear response function as is indicated in the figure with an S-shaped, sigmoid, curve. The function of a neuron can be described in mathematical form with:

$$Y=f(E)$$

$$E=\sum_{i=1}^{N} WiXi$$

where, Y is the output signal of the neuron and Xi are the input signals to the neuron, weighted with a factor Wi nonlinear function representing the threshold operation on the weighted sum of inputs.

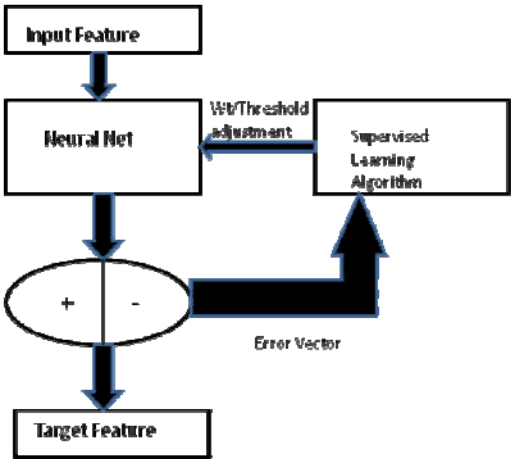
2. BRIDGE BETWEEN DATA MINING AND NEURAL NETWORKS

Neural-network methods are not commonly used for data-mining tasks, however, because they often produce incomprehensible models and require long training times.

Data Mining have several algorithms whose results are accurate and efficient, but not error prone. To deal with this kind of problems where Error detection and auto correction are needed, we can use Neural Networks as they have the capability of self error correction by adjusting weights as required. However the main problem of using this is long training time. Here in the next section we see what are the various training methods in Neural Networks.

3.TRAINING NEURAL NETWORKS

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 contain neural network.



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.

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 & a feedback from environment**.**

4. 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 manipulation 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. As data sets grow to massive sizes, the need for automated processing becomes clear. With their "model-free" estimators and their dual nature, neural networks serve data mining in a myriad of ways.

Data mining is the business of answering questions that you've not asked yet. Data mining reaches deep into databases. Data mining tasks can be classified into two categories: Descriptive and predictive data mining. Descriptive data mining provides information to understand what is happening inside the data without a predetermined idea. Predictive data mining allows the user to submit records with unknown field values, and the system will guess the unknown values based on previous patterns discovered form the database.

Data mining models can be categorized according to the tasks they perform: Classification and Prediction, Clustering, Association Rules. Classification and prediction is a predictive model, but clustering and association rules are descriptive models. The most common action in data mining is classification. It recognizes patterns that describe the group to which an item belongs. It does this by examining existing items that already have been classified and inferring a set of rules. Similar to classification is clustering. The major difference being that no groups have been predefined. Prediction is the construction and use of a model to assess the class of an unlabeled object or to assess the value or value ranges of a given object is likely to have. The next application is forecasting. This is different from predictions because it estimates the future value of continuous variables based on patterns within the data. Neural networks, depending on the architecture, provide associations, classifications, clusters, prediction and forecasting to the data mining industry. In data warehouses, neural networks are just one of the tools used in data mining. ANNs are used to find patterns in the data and to infer rules from them. Neural networks are useful in providing information on associations, classifications, clusters, and forecasting. The back propagation algorithm performs learning on a feed-forward neural network.

Feedforward Neural Network:

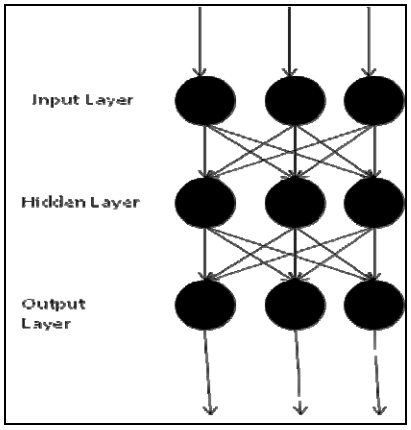


Figure 2 Feedforward Neural Network

One of the simplest feed forward neural networks (FFNN), such as in Figure 2, 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 previous layer. There are connections between the PEs in each layer that have a weight (parameter) associated with them. This weight is adjusted during training.

Information only travels in the forward direction through the network - there are no feedback loops.
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.

Back Propagation Algorithm:

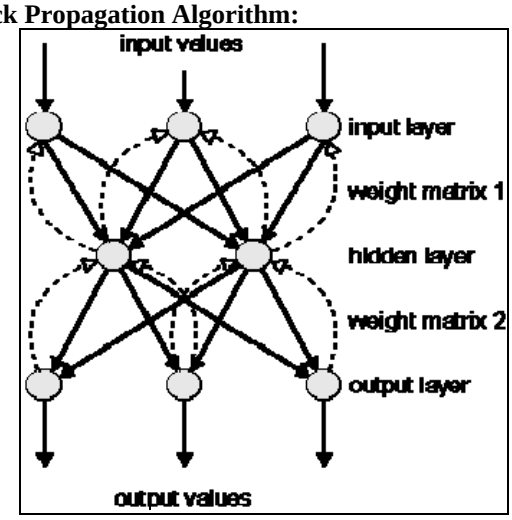


Figure 3 Back Propagation Neural Network

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.

Summary of the technique:

1. Present a training sample to the neural network.
2. Compare the network's output to the desired output from that sample. Calculate the error in each output neuron.
3. For each neuron, calculate what the output should have been, and a *scaling factor*, how much lower or higher the output must be adjusted to match the desired output. This is the local error.
4. Adjust the weights of each neuron to lower the local error.
5. Assign "blame" for the local error to neurons at the previous level, giving greater responsibility to neurons connected by stronger weights.
6. Repeat the steps above on the neurons at the previous level, using each one's "blame" as its error.

Actual Algorithm:

1. Initialize the weights in the network (often randomly)
2. repeat
* for each example e in the training set do
   1. O = neural-net-output(network, e) ; forward pass
   2. T = teacher output for e
   3. Calculate error (T - O) at the output units
   4. Compute delta_wi for all weights from hidden layer to output layer ; backward pass
   5. Compute delta_wi for all weights from input layer to hidden layer ;backward pass continued
   6. Update the weights in the network
* end
3. until all examples classified correctly or stopping criterion satisfied
4. return

5. Various Disciplines Where Data Mining Techniques Are Used Which Uses Neural Networks

There are numerous examples of commercial applications for neural networks. These include fraud detection, telecommunications, medicine, marketing, bankruptcy prediction, insurance, the list goes on. The following are examples of where neural networks have been used.

Accounting

- Identifying tax fraud
- Enhancing auditing by finding irregularities

Finance

- Signature and bank note verification
- Risk Management
- Foreign exchange rate forecasting
- Bankruptcy prediction
- Customer credit scoring
- Credit card approval and fraud detection
- Forecasting economic turning points
- Bond rating and trading
- Loan approvals
- Economic and financial forecasting

Marketing

- Classification of consumer spending pattern
- New product analysis
- Identification of customer characteristics
- Sale forecasts

Human resources

- Predicting employee's performance and behavior
- Determining personnel resource requirements

6. Advantages Of Neural Networks

1. High Accuracy: Neural networks are able to approximate complex non-linear mappings
2. Noise Tolerance: Neural networks are very flexible with respect to incomplete, missing and noisy data.
3. Independence from prior assumptions: Neural networks do not make a priori assumptions about the distribution of the data, or the form of interactions between factors.
4. Ease of maintenance: Neural networks can be updated with fresh data, making them useful for dynamic environments.
5. Neural networks can be implemented in parallel hardware
6. When an element of the neural network fails, it can continue without any problem by their parallel nature.

7. DESIGN PROBLEMS

- There are no general methods to determine the optimal number of neurons necessary for solving any problem.
- It is difficult to select a training data set which fully describes the problem to be solved.

8. CONCLUSION

There is rarely one right tool to use in data mining; it is a question as to what is available and what gives the "best" results. Many articles, in addition to those mentioned in this paper, consider neural networks to be a promising data mining tool. 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. Hence, the use of neural networks in data mining is a promising field of research especially given the ready availability of large mass of data sets and the reported ability of neural networks to detect and assimilate relationships between a large numbers of variables.

In most cases neural networks perform as well or better than the traditional statistical techniques to which they are compared. Resistance to using these "black boxes" is gradually diminishing as more statistical backgrounds. Thus, neural networks are becoming very popular with data mining practitioners, particularly in medical research, finance and marketing. This is because they have proven their predictive power through comparison with other statistical techniques using real data sets. Due to design problems neural systems need further research before they are widely accepted in industry. As software companies develop more sophisticated models with user-friendly interfaces the attraction to neural networks will continue to grow.

REFERENCES

[1] Hongjun Lu, Rudy Setiono, and Huan Liu, " Effective Data Mining Using Neural Networks," *IEEE Trans on Know and Data Engg*, Vol 8., No. 6, Dec 2003.

[2] NEURAL NETWORKS IN DATA MINING, " Journal 2009"

[3] Agrawal, R., Imielinski, T., Swami, A., "Database Mining: A Performance Perspective", *IEEE Transactions on Knowledge and Data Engineering*, pp. 914-925, December 2007 [3] Berry, J. A., Lindoff, G., Data Mining Techniques, Wiley Computer Publishing, 2001 (ISBN 0-471-17980-9).

[4] Effective Data Mining Through Neural Network,"IJ March-2012"

[5] Haykin, S., *Neural Networks*, Prentice Hall International Inc., 1999

[6] Berry, J. A., Lindoff, G., Data Mining Techniques, Wiley Computer Publishing, 1997 (ISBN 0-471-17980-9).

[7] Bhavani,Thura-is-ingham, "Data-mining Technologies,Techniques tools & Trends", CRC Press

[8] Bradley, I., Introduction to Neural Networks, Multinet Systems Pty Ltd 1997.

[9] Fayyad, Usama, Ramakrishna " Evolving Data mining into solutions for Insights", communications of the ACM 45, no. 8

[10] Fausett, Laurene (1994), Fundamentals of Neural Networks: Architectures, Algorithms and Applications, Prentice-Hall, New Jersey, USA.