A CONCISE ARTIFICIAL NEURAL NETWORK IN DATA MINING

Aleem Ali*

ABSTRACT

Data mining a multidisciplinary field, is an analytic process designed to explore data (typically business or market related) in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. There are many technologies available to data mining practitioners, including Artificial Neural Networks, Regression, and Decision Trees. Many practitioners are more concern to Neural Networks due to their black box nature, even though they have proven themselves in many situations. In our current research we are attempting to determine if Neural Networks outperform more traditional statistical techniques. This paper is an overview of artificial neural networks and questions their position as a preferred tool by data mining practitioners.

Keywords: *Artificial Neural Network (ANN), Data mining, Competitive learning, Multilayered feedforward neural network, Implementation, Advantages.*

*Jamia Hamdard, Hamdard University, Hamdard Nager, New Delhi, India.

1. INTRODUCTION

The term data mining is often used to apply to the two separate processes of knowledge discovery and prediction. Knowledge discovery provides explicit information that has a readable form and can be understood by a user. Forecasting, or predictive modeling provides predictions of future events and may be transparent and readable in some approaches (e.g. rule based systems) and opaque in others such as neural networks. Moreover, some data mining systems such as neural networks are inherently geared towards prediction and pattern recognition, rather than knowledge discovery**.** 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. Another example of where this institution has failed to utilize its data-warehouse is in cross-selling insurance products (e.g. home, life and motor vehicle insurance). By using transaction information they may have the ability to determine if a customer is making payments to another insurance broker. This would enable the institution to select prospects for their insurance products. These are simple examples of what could be achieved using data mining.

Four things are required to data-mine effectively: high-quality data, the "right" data, an adequate sample size and the right tool. There are many tools available to a data mining practitioner. These include decision trees, various types of regression and neural networks.

Income is a very important socio-economic indicator. If a bank knows a person's income, they can offer a higher credit card limit or determine if they are likely to want information on a home loan or managed investments. Even though this financial institution had the ability to determine a customer's income in two ways, from their credit card application, or through regular direct deposits into their bank account, they did not extract and utilize this information.

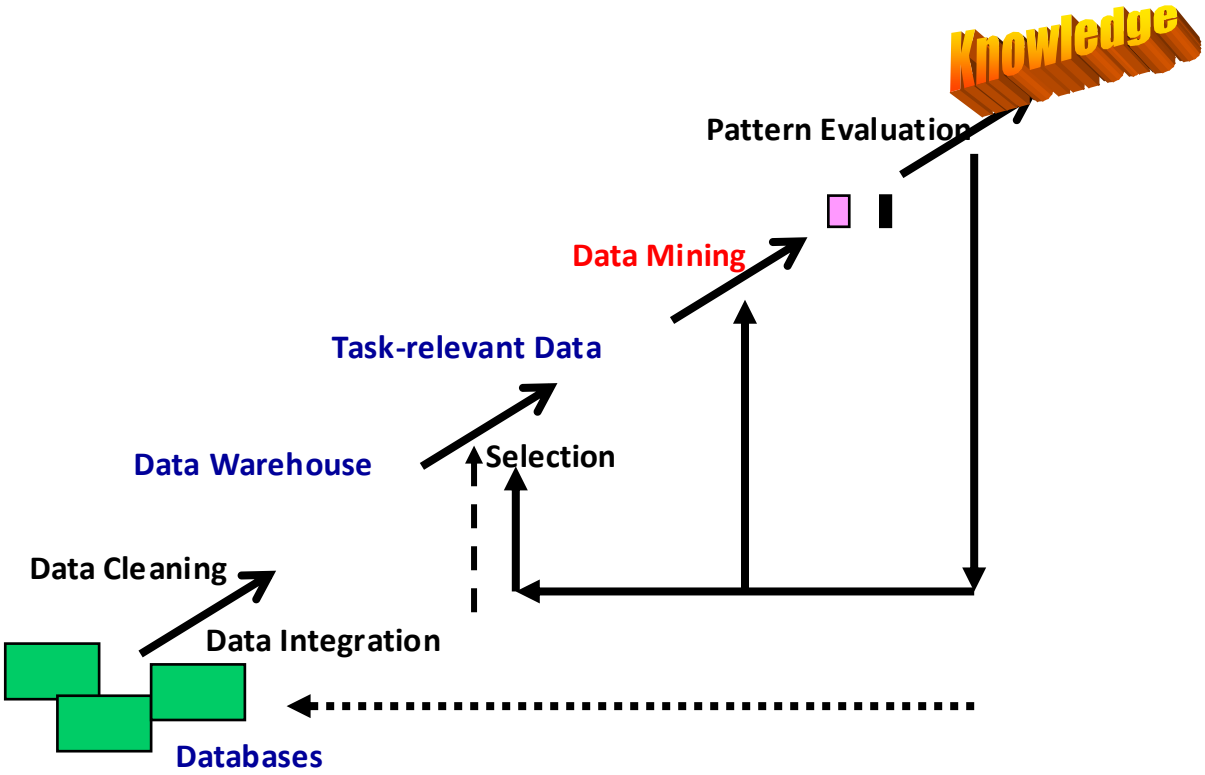

Figure 1: Operational Database Management Systems

2. METHODS OF DATA MINING

Classical methods of gaining knowledge from data sets are statistical methods. In data mining, new methods besides statistical are used. These methods have their origin in artificial intelligence. They differ from classical statistical methods in the following:

1. They look for unknown and unexpected relations, which can be uncovered by exploring data in database. They try to find such regularities in data sets, which would bring the user to the new view on his field of interest and allow him to formulate new hypotheses. Statistical methods are used in a rather different way. They verify or reject hypotheses stated a priori.
2. Data mining methods can be used also in such cases, in which utilization of classical statistical methods is not appropriate. For example, when a large volume of multivariate data is concerned or when it is not possible to suppose, that data have some standard probabilistic distribution.

In data mining, the following methods for gaining knowledge are studied and used.

– Statistical methods (prediction of time series, cluster analysis etc.)

– Production rules IF . . . THEN

– Decision trees

– Genetic algorithms

– Artificial Neural networks

2.1 Production rules IF . . . THEN

Production rules form the knowledge base of expert systems with production system architecture. In expert system design, the formulation of production rules are usually result of discussion between the knowledge en- gineer and a team of experts. In data mining, the methods of automatic formulation of production rules are studied. Such methods are mainly elaborated for production rules called association rules.

The purpose of association rules is to discover the associations among data in large databases, i.e., to find items that imply the presence of other items in the same transaction. Association rules were first introduced by Agraval et al. (1993).

Suppose $I = \{i_1, i_2, \ldots, i_n\}$ be a set of items. Let D be a set of transactions, where each transaction T is a set of items, such as $T \subseteq I$.

An association rule is an implication of the form $X \Rightarrow Y$, where

$X \subseteq I, Y \subseteq I, X \cap Y = \emptyset$

Association rule $X \Rightarrow Y$ holds in the database D with confidence c, if c% of transactions in D, that contain X, also contains Y.

The association rule $X \Rightarrow Y$ has in database D support

s, if s% transactions in D contain $X \cup Y$.

Mining association rules mean to find out all associa- tion rules that have support and confidence greater than or equal to the user specified minimum support (minsup) and minimum confidence (minconf).

Problems of automatic mining of association rules were intensively pursued during the last decade and effective algorithms were designed. See for example Agraval (1994) and Holt (2001).

2.2 Decision trees

Decision tree is a possible representation of a decision function. It is used when the complete knowledge of data is not necessary for appropriate decision and when the process of gaining data is expensive. Decision tree de- termines which data and in which order one should col- lect to achieve the effective decision with minimal average cost. Decision tree thus represents knowledge and can be used for effective decision-making. There exist algorithms for automatic construction of decision trees. Automatic construction of decision trees is the tra- ditional part of artificial intelligence.

2.3 Genetic algorithms

Genetic algorithm is a universal method for creating objects with desired properties (Holland 1975)). The method was inspired by Darwin's evolution theory. Objects are described by a sequence of symbols, which form the analogy of genom of living organisms. Capa- bility of an object to fulfil some function is measured by fitness function. Genetic algorithm creates generations of objects. New generation arises by mutation, cross- ing or by their combination from those objects of actual generation, which have the greatest value of the fitness function. Evolution modelled by genetic algorithm is aimed at the generation of objects with great value of fitness function or in other words, to the objects with the desired properties.

2.4 Artificial Neural Networks

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.

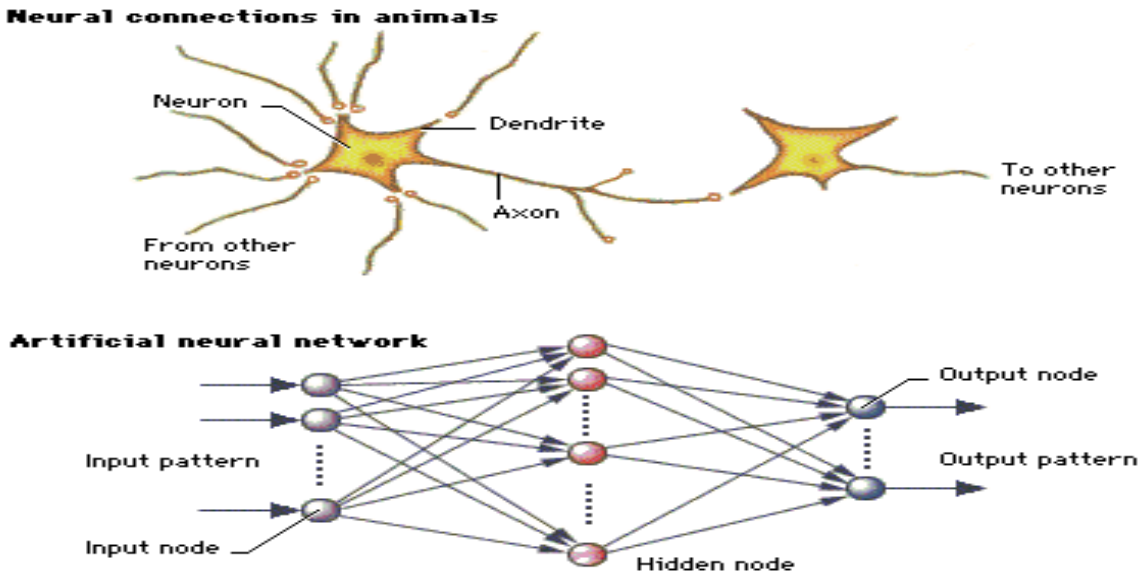

Figure 2: A typical neuron

3. TRAINING OF ARTIFICIAL NEURAL NETWORKS

An **artificial neural network** has to be configured such that the application of a set of inputs produces (either 'direct' or via a relaxation process) the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly,

using a priori knowledge. Another way is to **'train' the artificial neural network** by feeding it teaching patterns and letting it change its weights according to some learning rule. We can categorize the learning situations as follows:

3.1 Supervised learning or Associative learning in which the network is trained by providing it with input matching output patterns. These input-output pairs can be provided by an external teacher, or by the system which contains the neural network (self-supervised).

3.2 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. 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 of the input stimuli. Here 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 parameters.

4. ARTIFICIAL NEURAL NETWORK IN DATA MINING

Neural Networks are analytic techniques modeled after the (hypothesized) processes of learning in the cognitive system and the neurological functions of the brain and capable of predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a process of so-called *learning* from existing data. Neural Networks is one of the Data Mining techniques.

The first step is to design a specific network architecture (that includes a specific number of "layers" each consisting of a certain number of "neurons"). The size and structure of the network needs to match the nature (e.g., the formal complexity) of the investigated phenomenon. Because the latter is obviously not known very well at this early stage, this task is not easy and often involves multiple "trials and errors." (Now, there is, however, neural network software that applies artificial intelligence techniques to aid in that tedious task and finds "the best" network architecture.)

The new network is then subjected to the process of "training." In that phase, neurons apply an iterative process to the number of inputs (variables) to adjust the weights of the network in order to optimally predict (in traditional terms, we could say find a "fit" to) the sample data on which the "training" is performed. After the phase of learning from an existing data set, the new network is ready and it can then be used to generate predictions.

However, it should be mentioned that *Neural Network* techniques can also be used as a component of analyses designed to build explanatory models because *Neural Networks* can help explore data sets in search for relevant variables or groups of variables; the results of such explorations can then facilitate the process of model building. Moreover, now there is neural network software that uses sophisticated algorithms to search for the most relevant input variables, thus potentially contributing directly to the model building process.

4.1 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 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.

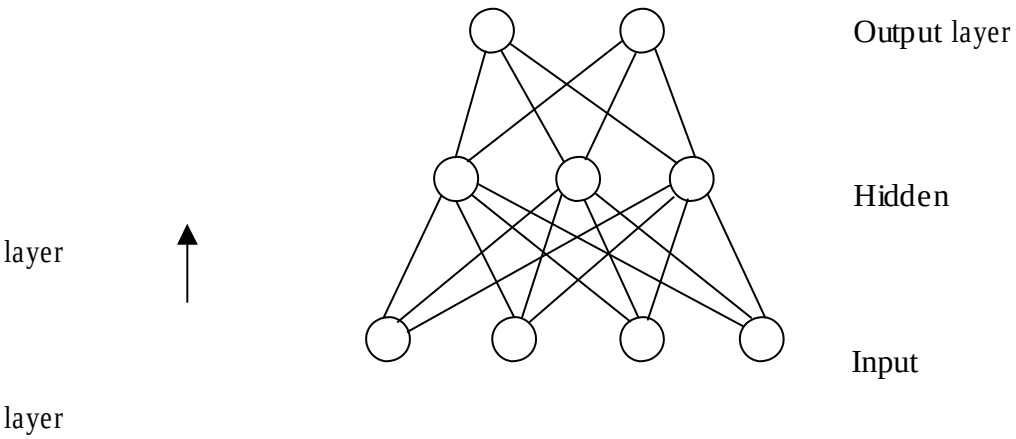

Figure 3: Multilayered feed-forward neural network (ANN)

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.

4.2 An Unsupervised method to trained the Artificial Neural Network

Unsupervised learning involves the use of inductive method to discover regularities that are present in the data set. Although there is a wide variety of neural network algorithms for unsupervised learning but competitive learning algorithm (Rumelhart & Zipser, 1985) is arguably the unsupervised neural network algorithm that is most appropriate for data mining, and it is illustrative of the utility of single layer neural network method.

The learning task addressed by competitive learning is to partition a given set of training examples into a finite set of clusters. The clusters should represent regularities present in the data set such that similar example are mapped into similar classes.

The variant of competitive learning that we consider here, which is sometimes called *simple competitive learning*, involves learning in a single layer network. The input units in such a network represent the relevant features of the problem domain, and k output units represent the k classes into which examples are clustered.

The net input to each output unit in this method is a linear combination of the input activation:

$$net_i = \sum_j w_{ij}\, x_j .$$

Here, x_j is the activation of the j^{th} input unit, and w_{ij} is the weight linking the jth input unit to the i^{th} output. The name *competitive learning derives* from the process used to determine the activations of the hidden units. The output unit that has the greatest net input is deemed the winner, and its activation is set to one. The activations of the other output are set to zero:

$$x_i = \begin{cases} 1 & \text{if } \sum_j w_{ij} x_j > \sum_j w_{hj} x_j \text{ for all output units } h \neq i \\ 0 & \text{Otherwise} \end{cases}$$

The training process for *competitive learning* involves minimizing the cost function:

$$C = \tfrac{1}{2} \sum\sum x_i (x_j - w_{ij})^2$$

i j

Where x_i is the activation of the i^{th} output unit, x_j is the activation of the j^{th} input unit, and w_{ij} is the weight from the j^{th} input unit to the i^{th} output unit. The update rule for the weight is then:

$\Delta w_{ij} = -\eta \partial C \partial w_{ij} = \eta x_i(x_j - wij)$.

Where η learning rate parameter.

The basic idea competitive learning is that each output unit takes responsibility for a subset of the training examples. Only one output unit is the winner for a given example, and the weight vector for the winning unit is moved towards the input vector for this example. As training progresses, therefore, the weight vector of each output unit moves towards the centroid of the examples for which the output unit has taken responsibility. After training each output unit represents the cluster of examples, and the weight vector for the unit corresponds to the centroid of the cluster.

Competitive learning is closely related to the statistical method known as *k-means clustering*. The principal difference between the two methods is that competitive learning is an *online algorithm*, meaning that during training it updates the network's weight after every example is presented, instead of after all of the examples have been presented. The online nature of competitive learning makes it more suitable for very large data sets, since online algorithm usually converge to a solution faster in such cases.

5. ADVANTAGES OF ARTIFICIAL NEURAL NETWORK

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.

6. DESIGN PROBLEMS

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

7. CONCLUSION

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. The combination of data mining method and neural network model can greatly improve the efficiency of data mining methods, and it has been widely used. It also will receive more and more attention.

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 researchers use them, in particular those with 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. Fausett, Laurene (1994), Fundamentals of Neural Networks: Architectures, Algorithms and Applications, Prentice-Hall, New Jersey, USA.
2. Agrawal R, Srikant R. (1994): Fast algorithms for mining association rules. Proceedings of the 20th VLDB Confer- ence: 487-499.
3. Data Mining:Practical Machine Learning Ian H. Witten, Eibe Frank, Morgan Kaufmann, 2000.
4. http://en.wikipedia.org
5. Berry, J. A., Lindoff, G., *Data Mining Technique,* , Wiley Computer Publishing, 1997
6. (ISBN 0-471-17980-9).
7. Edelstein, H. (1996) "Mining data warehouses," *InformationWeek*, Jan 08, **561**, p. 48(4).

8. Rumelhart, D. E & Zipser, D. (1985). Feature discovery by competitive learning. Cognitive Science, 9:75-112.
9. Zurada J.M., "An introduction to artificial neural systems", St. Paul: West Publishing 1992).
10. Adriaans P, Zantinge D. Data mining [M]. Addision_Wesley Longman, 1996.