

Artificial Neural Networks in Data Mining

Nashaat El-Khamisy Mohamed^{#1}, Ahmed Shawky Morsi El-Bhrawy^{*2}

^{#*} (Computer & Information System, Sadat Academy for Management sciences, Egypt)

Abstract: There are plenty of technologies available to data mining practitioners, including Artificial Neural Networks, Regression, and Decision Trees. Data mining tools can forecast the future trends and activities to support the decision of Officials. Some data mining tools can also solve some traditional problems which consumed enough time, this is because that they can rapidly browse the entire database and find some useful information experts undetected. Neural network is a parallel processing network which made with simulating the intuitive thinking of human, the neural network in data mining was not optimistic, and the key reasons are that the nerve organs network has the problems of complex structure, poor interpretability and long training time. Nevertheless advantages such as high affordability to the noise data and low error rate, the consistently advancing and optimization of various network training algorithms, especially the continually advancing and improvement of varied network pruning algorithms and rules extracting algorithm, associated with application of the neural network in the data mining increasingly preferred by the overwhelming vast majority of users. This paper introduces the Comprehensive view of artificial neural networks and their advantages by data mining practitioners.

Keyword s: Artificial Neural Network (ANN), neural network topology, Data mining, back propagation algorithm, Advantages.

I. Introduction

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 kind 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 about what they deem to be important offer a higher credit card limit or determine if they are likely to want information on a home loan or managed investments. Also though this financial institution experienced the ability to determine a customer's income in two different ways, from their credit card program, or through regular immediate deposits into their standard bank account, they did not extract and utilize this information. Another example of where this institution has failed to utilize their data-warehouse is in cross-selling insurance products (e. g. home, life and motor vehicle insurance). By using transaction information they could have the ability to determine if a customer is making payments to another insurance professional. This would permit the institution to select prospects for their insurance products. These are simple types of what could be achieved using data mining. Four things are necessary to data mine effectively: top quality data, the "right" data, an enough sample size and the right tool. There are many tools available to a data mining specialist. These include decision trees, various types of regression and neural networks [1].

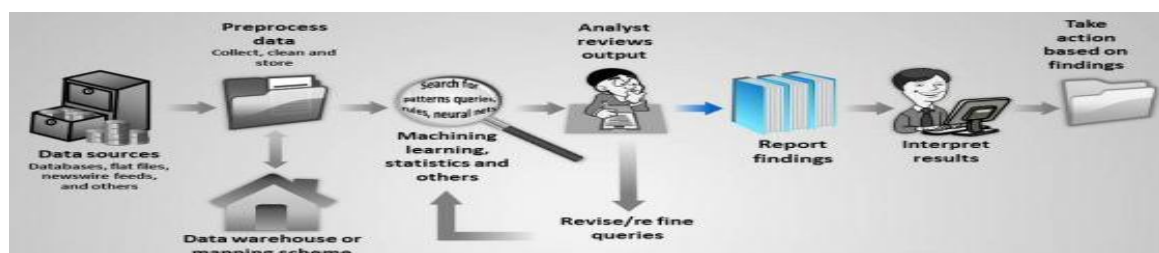


Fig.1 Data mining process

II. 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. 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. A neural network is an interconnected group of nodes, akin to the vast network of neurons in the human brain [2].

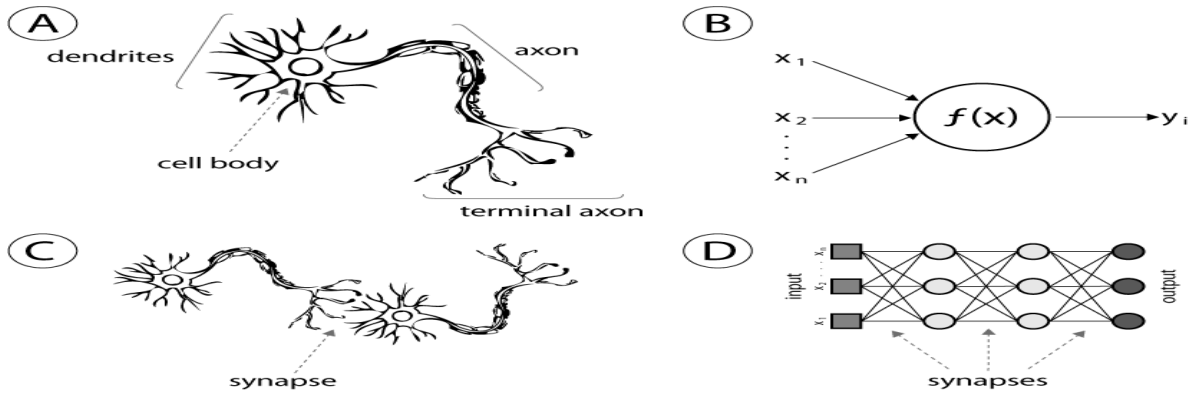


Fig.2 (A) Human neuron; (B) artificial neuron or hidden unity; (C) biological synapse; (D) ANN synapses

III. Training of Artificial Neural Networks

A 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 [3]. 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 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:

- **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 contains the neural network (self-supervised) [4].

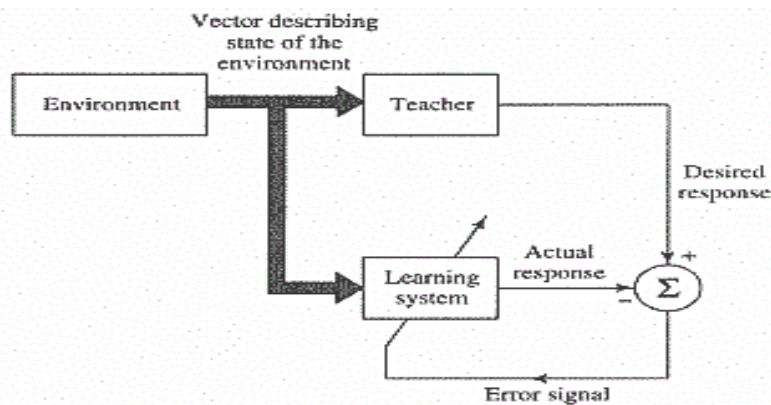


Fig.3 Neural Network Supervised Learning

- **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 [5].

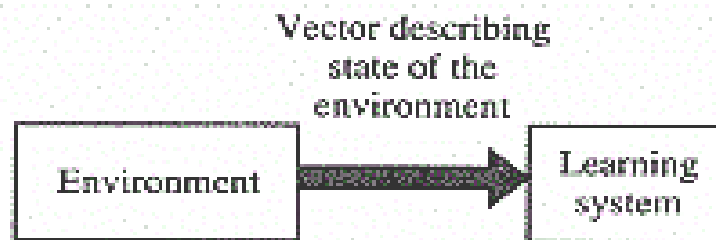


Fig.4 Neural Network Unsupervised Learning

• 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 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 [5].

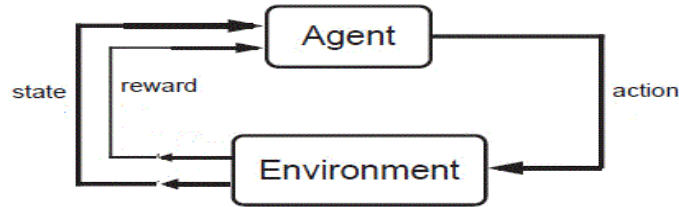


Fig.5 Neural Network Reinforcement Learning

IV. Neural Network Techniques

A. 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 [7].

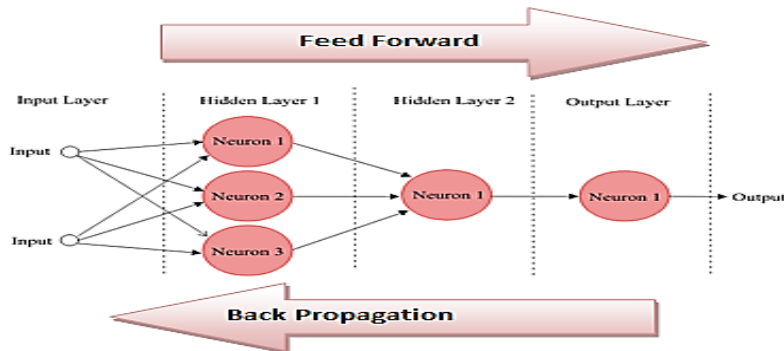


Fig.6 Neural Network Techniques

A PE receives inputs from either the outside world or the simplified process for training a FFNN is as follows [8]:

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.

B. The Back Propagation Algorithm

Back propagation, 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 feed-forward 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 [2].

V. 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.

VI. 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 (network)

VII. Review of Literature Reporting Neural Network Performance

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 [6], [7].

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

VIII. The Advantages And Disadvantages of Neural Networks

Table I the Advantages and Disadvantages of Neural Networks

Advantages of Neural Networks	Disadvantages of Neural Networks
High Accuracy: Neural networks are able to approximate complex non-linear mappings.	Poor Transparency: Neural networks operate as "black boxes".
Noise Tolerance: Neural networks are very flexible with respect to, incomplete, missing and noisy data.	Trial-and-error design: The selection of the hidden nodes and training parameters is heuristic.
Independence from prior assumptions: Neural networks do not make prior assumptions about the distribution of the data, or the form of interactions between factors.	Data hungry: Estimating the network weights requires large amounts of data, and this can be very computationally intensive.
Ease of maintenance: Neural networks can be updated with fresh data, making them useful for dynamic environments.	Over-fitting: If too many weights are used without regularization, Neural network become useless in terms of generalization to new data.
Neural network overcomes some limitations of other statistical methods while generalizing them.	There is no explicit set of rules to select the most suitable Neural network algorithm.
Hidden nodes, in supervised Neural networks can be regarded as latent variables.	Neural networks are totally dependent on the quality and amount of data available.
Neural networks can be implemented in parallel hardware.	Neural networks may converge to a local minimum in the error surface.
Neural network performance can be highly automated, minimizing human involvement.	Neural networks lack classical statistical properties. Confidence intervals and hypothesis testing are not available.
Neural networks are especially suited to tackling problems in non-conservative domains.	Neural network techniques are still rapidly evolving and they are not yet robust

Design Problems

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

Solutions to Improve Ann performance

- Designing Neural Networks using Genetic Algorithms
- Neuro-Fuzzy Systems

IX. Conclusion

There is certainly rarely one right tool to use in data mining; this can be a question as to what is available and what gives the "best" results. Many articles, in conjunction with 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 credited to the complexity in translating the systems into precise mathematical functions. Therefore, the use of neural networks in data exploration is a promising field of research especially given the ready availability of large mass of information sets and the reported ability of neural networks to discover and absorb relationships between sizable quantities of variables. In most cases neural networks perform as well or better than the regular statistical techniques to which they are compared. Resistance to using these "black boxes" is slowly but surely diminishing as more researchers use them, in particular those with statistical backgrounds. Thus, neural networks becoming very popular with data mining experts, particularly in medical research, finance and marketing. This kind of 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 complex models with user-friendly interfaces the attraction to neural networks will continue to grow.

References

- [1] M.Charles Arockiaraj , "Applications of Neural Networks In Data Mining", International Journal Of Engineering And Science, Vol.3, Issue 1 (May 2013), PP 08-11
- [2] Sonalkadu, Sheetal Dhande "Effective Data Mining Through Neural Network" , International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 2, Issue 3, March 2012 ISSN: 2277 128X
- [3] J.Rabunal, J.Dorado, "Artificial neural networks in real-life applications", 2006, p 297-303.
- [4] Farhad Bilal Baha'addin,"Kurdistan Engineering Colleges and Using of Artificial Neural Network for Knowledge Representation in learning process", (IJEIT) Volume 3, Issue 6, December 2013, p 292:299.
- [5] Mohamed M. Mostafa,"Profiling blood donors in Egypt: A neural network analysis", Volume 36, Issue 3, Part 1, April 2009, p 5031-5038.
- [6] Nainja Rikhi ,"Data Mining and Knowledge Discovery in Database ", International Journal of Engineering Trends and Technology Volume23 Number 2, May 2015.
- [7] S.Xu, M.Zhang, "Data mining - an adaptive neural network model for financial analysis" Information Technology and Applications, IEEE, ICITA, 2005, p 336-340.
- [8] X.Ni, "Research of Data Mining Based on Neural Networks", World Academy of Science, Engineering and Technology 39, 2008.