

Neural Networks in Data Processing

Ajay Gupta

University Of Mumbai(MCA)

ASM's Institute of Management & Computer Studies

C-4, Wagle Industrial Estate, Check Naka, Thane(W)-400604

ABSTRACT--The application of neural networks within the data processing is incredibly wide, though neural Networks could have complicated structure, long training time, and apprehensively comprehensible illustration of results, neural networks have high acceptance ability for noise acceptance and high accuracy and are preferred in data processing. In this paper the data mining supported neural networks is researched thoroughly, and therefore the key technology and ways that to realize the data mining supported neural networks are researched.

KEYWORDS-- Data processing, Neural networks, Artificial neural network (ANN), Data mining method, Implementation.

1. INTRODUCTION

Data mining refers to method or technique that extracts or mines fascinating information or patterns from massive amounts of information. It searches for relationship and patterns that exist in large database however area unithidden among large quantity of information. It is referred as information dredging. Data processing technique are:

1. Classification: it's organization of information in given category. It uses a training set where all the objects are already associated with known category label. It is also referred as supervised classification.

2. Clustering: It refers to grouping of records. Observations into categories of similar objects. It's referred as unsupervised classification, it is based on principle of maximizing similarity between objects in same category and minimizing similarity between objects of various category.

3. Association: It studies frequency of data items occurring along in transactional database. Support identifies frequent item set. Confidence is

condition likelihood that an item appear in transaction when another item appears.

2. DATA MINING METHOD

In data processing numerous ways are used for gaining knowledge Such as:

1. Decision tree: Its hierarchical cluster of relationship organized into tree like structure. It determines that knowledge and within which order should be collected to realize effective Decision that represent data with nominal cost.

2. Genetic algorithm: They are category of randomized search procedure capable of adjective search over a wide range of search space topology. They possess the ability to solve problem in parallel, therefore it's a robust tool for data processing. Genetic algorithmic rule search uses a collection of solutions throughout every generation.

3. Neural network: It is directed graph consisting of nodes with interconnecting synaptic and activation link. It's a data processing network that is generated by stimulating image intuitive thinking of human on basis of analysis of biological neural network. It is additionally referred as Artificial Neural Network. An Artificial Neural Network (ANN) is impressed by the ways biological nervous systems, like the brain, process data.

3. NEURAL NETWORK METHOD IN DATA MINING

In neural network knowledge representation of surrounding environment is outlined by value taken on by parameters like synaptic weight and biases of network. Three neural network methodology.

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

b) **Feedback network:** It regards perception Hopfield separate model and continuous Model as representative.

c) **Self organization**

network: It regards adjective Resonance theory and Kohen model as representative Pattern recognition and performance estimation talents of neural network square measure utilized in data processing. Neural network was trained to store, acknowledge and retrieve pattern or info entries, to filter noise from measurement data.

4. ARTIFICIAL NEURAL NETWORKS:

It is a computational system inspired by the Structure Processing Method Learning Ability of a biological brain .An artificial neural network (ANN), typically simply known as a "neural network" (NN), may be a mathematical model or process model based on biological neural networks, in different words, is an emulation of biological neural system. It consists of interconnected cluster of artificial neurons and processes information using a connectionist approach to computation.

4.1 Neural Network Topologies:

- **Feed forward neural network:** The feedforward neural network was the primary and arguably simplest form of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There aren't any cycles or loops within the network. The data processing can extend over multiple (layers of) units, however no feedback connections are present, that is, connections extending from outputs of units to inputs of units within the same layer or previous layers.

- **Recurrent network:** Recurrent neural networks are those that can contain feedback connections. Contrary to feed forward networks, recurrent neural networks (RNs) are models with bi-directional information flow. Whereas a feedforward network propagates information linearly from input to output, RNs also propagate information from later processing stages to earlier stages.

4.2 Training of Artificial Neural Networks:

Once a network has been structured for a particular application, that network is ready to be trained. To start this process the initial weights are chosen randomly. Then, the training, or learning, begins. A neural network has got to be designed specified the appliance of a collection of inputs produces (either 'direct' or via a relaxation process) the required set of outputs. Numerous ways to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge ,in a different way is to 'train' the neural network by feeding it teaching patterns and letting it modify its weights in keeping with some learning rule .

We can categorize the learning situations as follows:

- **Supervised learning** or Associative learning within which the network is trained by providing it with input and matching output patterns. These input-output pairs can be provided by associate external teacher, or by the system that contains the neural network (self-supervised).
- **Unsupervised learning** or selforganization within which an (output) unit is trained to respond to clusters of pattern within the input. During this paradigm the system is meant to find statistically salient features of the input population. Not like the supervised learning paradigm, there's not a priori set of categories into which the patterns are to be classified; rather the system should develop its own illustration of the input stimuli.

- **Reinforcement learning** this sort of learning is also thought of as associate intermediate form of the above two forms of learning. Here the learning machine does some action on the environment and gets a feedback response from the environment, system grades its action good (rewarding) or bad (punishable) based on the environmental response and consequently adjusts its parameters.

5. DATA MINING PROCESS BASED ON NEURAL NETWORK

1. . **Data preparing:** It defines and process mining data to make it work specific data processing methodology. It includes:

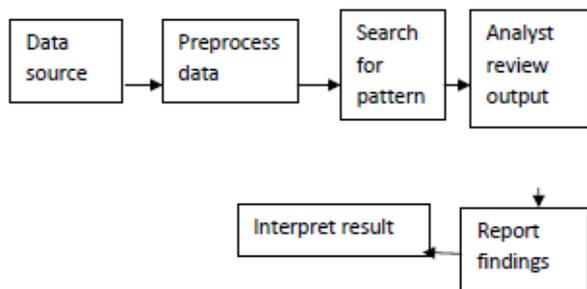
a) **Data cleaning:** It routine work to clean data by filling in missing values, smoothing noise data, removing outliers and resolve inconsistency.

b) **Data option:** it is used to select arrangement of data,

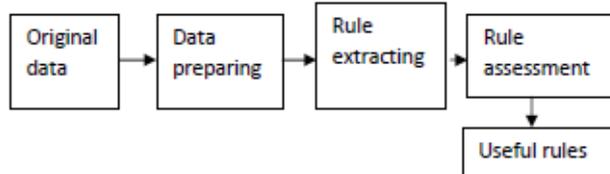
c) **Data preprocessing:** It describes any form of process performed on raw data to organize it for one more processing procedure.

d) **Data transformation:** here selected data is transformed into a type applicable for mining procedure. Data processing supported neural network will handle numeric data only.

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 numerous methods for extracting rules. These are mainly

a) LRE methodology

b) Black Box methodology

c) BIO-RE

3. Rule assessment:

It is basically done for finding optimum sequence of extracting rules, check accuracy of rules extracted and to observe what proportion information in neural network had not been extracted.

6. NEURAL NETWORK LEARNING ALGORITHM

Back propagation may be a common methodology of training artificial neural networks so as to minimize the objective function...it's a supervised learning methodology, and may be a generalization of the delta rule. It needs a dataset of the required output for several inputs, creating up the training set. It's most helpful for feedforward networks. The term is an abbreviation for "backward propagation of errors". Back propagation needs that the activation function used by the artificial neurons be differentiable... the. It 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 so as to come up with the propagation's output activations.

2. Backward propagation of the propagation's output activations through the neural network using the training pattern's target so as to come up with 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 so as to come up with 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 acquire 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's referred to as the *learning* rate. The sign of the gradient of a weight indicates wherever the error is increasing; this is why the weight should be updated within opposite direction. Back propagation is a process that may usually take a good deal of time to complete.

7. REVIEW OF LITERATURE COVERAGE NEURAL NETWORK PERFORMANCE:

There various examples of commercial applications for neural networks. These include; fraud detection, telecommunications, medicine, marketing, bankruptcy prediction, insurance, the list goes on. The following examples of wherever neural networks are used.
Accounting Identifying tax fraud Enhancing auditing by finding irregularities
Finance Signature and bank note verification Risk Management Bankruptcy prediction Customer credit _ rating Credit card approval and fraud detection Bond rating and commerce Loan approvals Economic and money statement
Marketing Classification of client defrayal pattern New product analysis Identification of client characteristics Sale forecasts
Human resources Predicting employee's performance and behavior Determining personnel resource needs

8. ADVANTAGES OF NEURAL NETWORKS:

1. High Accuracy: Neural networks are able to approximate complicated non-linear mappings
2. Noise Tolerance: Neural networks are Flexible with relevance to incomplete, missing and noisy information.
3. Independence from previous assumptions: Neural networks don't build a priori assumptions regarding the distribution of the information, or the form of interactions between factors.
4. Easy maintenance: Neural networks can be updated with fresh information, creating them helpful for dynamic

environments.

5. Neural networks can be implemented in parallel hardware
6. Once a component of the neural network fails, it will continue with none downside by their parallel nature

9. 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.

10. SOLUTIONS TO IMPROVE ANN PERFORMANCE:

1. Designing Neural Networks using Genetic Algorithms
2. Neuro-Fuzzy Systems

11. CONCLUSION:

There is seldom one right tool to use in knowledge mining; it's a question on what's obtainable and what offers the "best" results. Several articles, additionally to those mentioned during this paper, take into account neural networks to be a promising data processing tool. Artificial Neural Networks provide qualitative methods for business and economic systems that ancient quantitative tools in statistics and economic science cannot quantify owing to the quality in translating the systems into precise mathematical functions. Hence, the utilization of neural networks in data processing may be a promising field of analysis particularly given the prepared handiness of enormous mass of data sets and therefore the reported ability of neural networks to find and assimilate relationships between an outsized numbers of variables. In most cases neural networks perform as well or better than the normal applied math techniques to that they're compared. Resistance to using these "black boxes" is step by step decreasing as additional researchers use them, particularly those with applied math backgrounds. Thus, neural networks have become

extremely popular with data processing practitioners, significantly in medical analysis, finance and selling. As software system firms develop sophisticated models with easy interfaces the attraction to neural networks can still grow. This paper presents analysis on data processing supported neural network. Neural network solves the matter of data mining because it has parallel processing, distributed storage, high degree of fault tolerance, good robustness. In this paper data {processing} process based on neural network, backpropagation algorithm are explained and emphasis is laid on numerous techniques for improving the performance like Using Genetic Algorithm with Neural Network and neuro fuzzy system.

12. ACKNOWLEDGEMENT

We would like to take this opportunity to express my profound gratitude and deep regard to our Guide Mrs. Reeta Singh, for her exemplary guidance, valuable feedback and constant encouragement throughout the duration of the research. Her valuable suggestions were of immense help throughout our research work. Her perceptive criticism kept us working to make this research in a much better way. Working under her was an extremely knowledgeable experience for us.

We would also like to give my sincere gratitude to all the friends and colleagues who filled in the survey, without which this research would be incomplete.

13. REFERENCES

- [1] Agrawal, R., Imielinski, T., Swami, A., "Database Mining: A Performance Perspective", *IEEE Transactions on Knowledge and Data Engineering*, pp. 914- 925, December 1993
- [2] Berry, J. A., Lindoff, G., Data Mining Techniques, Wiley Computer Publishing, 1997 (ISBN 0-471-17980-9).
- [3] Berson, "Data Warehousing, Data-Mining & OLAP", TMH
- [4] Bhavani, Thura-is-ingham, "Data-mining Technologies, Techniques tools & Trends", CRC Press
- [5] Bradley, I., Introduction to Neural Networks, Multinet Systems Pty Ltd 1997.
- [6] Fayyad, Usama, Ramakrishna "Evolving Data mining into solutions for Insights", communications of the ACM 45, no. 8

- [7] Fausett, Laurene (1994), Fundamentals of Neural Networks: Architectures, Algorithms and Applications, Prentice-Hall, New Jersey, USA.
- [8] Haykin, S., *Neural Networks*, Prentice Hall International Inc., 1999
- [9] Khajanchi, Amit, Artificial Neural Networks: The next intelligence
- [10] Zurada J.M., "An introduction to artificial neural networks systems", St. Paul: West Publishing (1992)