Data Mining Using Neural Networks

Neha Suyal
HP-India, Bangalore
suyal.neha1@gmail.com

Abstract- ***In the last few years the development of algorithm for knowledge discovery in data base (KDD) has been growing very fast. This technology involved the process of finding and interpreting patterns from huge and complex databases. The data mining phase of this technology describes efficient and intelligent technique that handle such data analyzing. In this paper I will discuss the data mining issues with the help of neural networks.***

Keywords- ***Data mining; KDD; Neural Network (NN); Neural Network Topology.***

I. INTRODUCTION

Data mining is a component of a wider process called knowledge discovery from database. It involves scientist from a wide range of disciplines, including mathematicians, computer scientists and statisticians as well as those working in the fields such as machine learning, artificial intelligence, information retrieval and pattern recognition.

A data warehouse is a subject-oriented, integrated, time-varying, non-volatile collection of data in support of the management's decision-making process. 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". Business enterprises are beginning to realize that information on customers and their buying patterns are most valuable information. Competitiveness increasingly depends on the quality of decision making and learning to make quality decision from past transactions and decisions. The large databases maintained by retailers, telecom service providers, and credit card companies contain very valuable Information related to customers. These enterprises could benefit immensely in the areas of marketing, advertising and sales, if interesting and previously unknown customer buying and calling patterns can be
discovered from the large volumes of data. 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.

II. NEURAL NETWORKS

Neural networks or Artificial Neural Networks (ANN) are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements.

Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. Such a situation is shown in fig (1). There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically many such input/target pairs are used, in this supervised learning, to train a network. Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision and control systems.

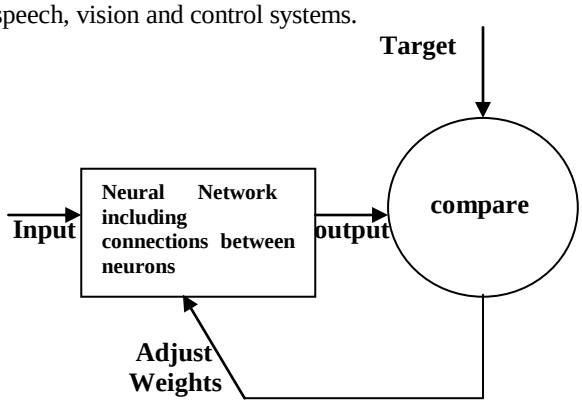

Fig. 1. : Neural Netwoks block diagram

Today neural networks can be trained to solve problems that are difficult for conventional computers or human beings. The supervised training methods are commonly used, but other networks can be obtained from unsupervised training techniques or from direct design methods. Unsupervised networks can be used, for instance, to identify groups of data. Certain kinds of linear networks and Hopfield networks are designed directly. In summary, there are a variety of kinds of design and learning techniques that enrich the choices that a user can make.

The field of neural networks has a history of some five decades but has found solid application only in the past fifteen years, and the field is still developing rapidly. Thus, it is distinctly different from the fields of control systems or optimization where the terminology, basic mathematics, and design procedures have been firmly

established and applied for many years.

There are two modes of learning: Supervised and unsupervised. Below there is a brief description of each one to determine the best one for our problem

A. Supervised Learning

Supervised learning is based on the system trying to predict outcomes for known examples and is a commonly used training method. It compares its predictions to the target answer and "learns" from its mistakes.

The data start as inputs to the input layer neurons. The neurons pass the inputs along to the next nodes. As inputs are passed along, the weighting, or connection, is applied and when the inputs reach the next node, the weightings are summed and either intensified or weakened.

This continues until the data reach the output layer where the model predicts an outcome. In a supervised learning system, the predicted output is compared to the actual output for that case. If the predicted output is equal to the actual output, no change is made to the weights in the system. But, if the predicted output is higher or lower than the actual outcome in the data, the error is propagated back through the system and the weights are adjusted accordingly. This feeding error backwards through the network is called "back-propagation." Both the Multi-Layer Perceptron and the Radial Basis Functions are supervised learning techniques. The Multi-Layer Perceptron uses the back-propagation while the Radial Basis Function is a feed-forward approach which trains on a single pass of the data.

B. Unsupervised Learning

Neural networks which use unsupervised learning are most effective for describing data rather than predicting it. The neural network is not shown any outputs or answers as a part of the training process--in fact, there is no concept of output fields in this type of system. The primary unsupervised technique is the Kohonen network. The main uses of Kohonen and other unsupervised neural systems are in cluster analysis where the goal is to group "like" cases together.

The advantage of the neural network for this type of analysis is that it requires no initial assumptions about what constitutes a group or how many groups there are. The system starts with a clean slate and is not biased about which factors should be most important.

III. NEURAL NETWORKS IN DATA MINING

Data mining refers to extracting or mining knowledge from large amounts of data. Many people treat data mining as a synonym for another popularly used term, knowledge discovery from data or KDD. Knowledge discovery consists of an iterative sequence of the following steps-

- **Data cleaning** –(to remove noise and inconsistent data)
- **Data integration** –(where multiple data sources may be combined)
- **Data selection** -(where data relevant to the analysis task are retrieved from the data base)
- **Data transformation** -(where data trans formed or consolidated into forms appropriate for mining by performing summary or aggregation operations for instance)
- **Data mining** – (an essential process where intelligent methods are applied in order to extract data patterns)
- **Pattern evaluation**- (to identify the truly interesting patterns representing knowledge based on some interestingness measures)
- **Knowledge Presentation** -(where visualization and knowledge representation techniques are used to present the mined knowledge to the user).

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. Financial forecasting is of considerable practical interest. Due to neural networks can mine valuable information from a mass of history information and be efficiently used in financial areas, so the applications of neural networks to financial forecasting have been very popular over the last few years. Some researches show that neural networks performed better than conventional statistical approaches in financial forecasting and are an excellent data-mining tool. 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.

IV. ADVANTAGES OF NEURAL NETWORKS

- High Accuracy: Neural networks are able to approximate complex non-linear mappings
- Noise Tolerance: Neural networks are very flexible with respect to incomplete, missing and noisy data.
- 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.
- Ease of maintenance: Neural networks can be updated with fresh data, making them useful for dynamic environments.
- Neural networks can be implemented in parallel hardware
- When an element of the neural network fails, it can continue without any problem by their parallel nature.

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

V. 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 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] 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 vol 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] Zurada J.M., "An introduction to artificial neural networks systems", St. Paul: West Publishing (1992)