

Neural Networks in Data Mining

Priyanka Gaur

Department of Information and Technology
Marudhar Engineering College, Bikaner, Rajasthan
[Email- er.priyankagaur@gmail.com](mailto:er.priyankagaur@gmail.com)

Abstract: The application of neural networks in the data mining is very wide. Although neural networks may have complex structure, long training time, and uneasily understandable representation of results, neural networks have high acceptance ability for noisy data and high accuracy and are preferable in data mining. In this paper the data mining based on neural networks is researched in detail, and the key technology and ways to achieve the data mining based on neural networks are also researched.

Keywords – Data mining, neural networks, artificial neural network (ANN), data mining process, implementation.

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 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”.

Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural networks or decision trees). Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction.

A number of advances in technology and business processes have contributed to a growing interest in data mining in both the public and private sectors. Some of these changes include the growth of computer networks, which can be used to connect databases; the development of enhanced search-related techniques such as neural networks and advanced algorithms; the spread of the client/server computing model, allowing users to access centralized data resources from the desktop; and an increased ability to combine data from disparate sources into a single search source.

II. NEURAL NETWORK

Neural networks represent a brain metaphor for information processing. These models are biologically inspired rather than an exact replica of how the brain actually functions. Neural networks have been shown to be very promising systems in many forecasting applications and business classification applications due to their ability to “learn” from the data, their nonparametric nature (i.e., no rigid assumptions), and their ability to generalize.

Neural computing refers to a pattern recognition methodology for machine learning. The resulting model from neural computing is often called an **artificial neural network (ANN)** or a *neural network*. Neural networks have been used in many business applications for pattern recognition, forecasting, prediction, and classification. Neural network computing is a key component of any data mining tool kit.

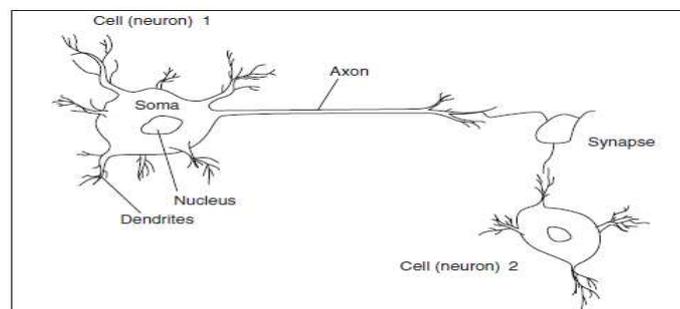


Figure 1. Two interconnected biological cells.

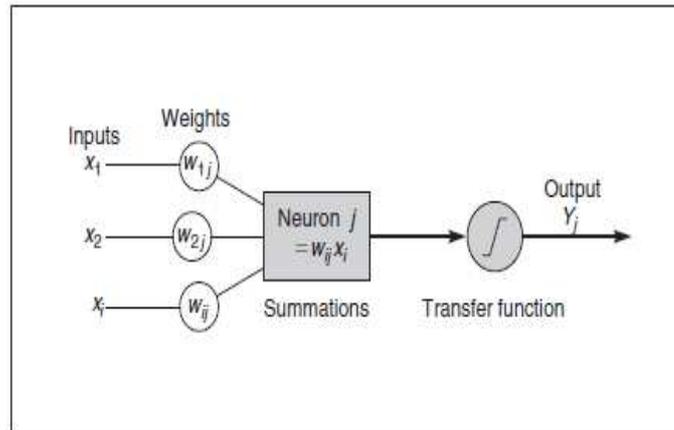


Figure 2. Processing in an Artificial Neuron.

A. NEURAL NETWORK METHOD IN DATA MINING

Neural network method is used for classification, clustering, feature mining, prediction and pattern recognition. It imitates the neurons structure of animals, bases on the M-P model and Hebb learning rule, so in essence it is a distributed matrix structure. Through training data mining, the neural network method gradually calculates (including repeated iteration or cumulative calculation) the weights the neural network connected. The neural network model can be broadly divided into the following three types:

- (a) *Feed-forward networks*: It regards the perception back-propagation model and the function network as representatives, and mainly used in the areas such as prediction and pattern recognition;
- (b) *Feedback network*: It regards Hopfield discrete model and continuous model as representatives, and mainly used for associative memory and optimization calculation;
- (c) *Self-organization networks*: it regards adaptive resonance theory (ART) model and Kohonen model as representatives, and mainly used for cluster analysis.

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 anipulation 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.

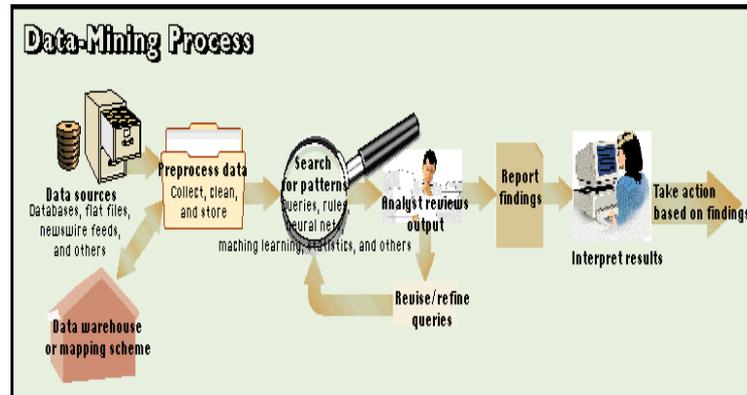


Figure 3. Image of data-mining process.

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 from 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.

- i. *Classification*: 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.
- ii. *Clustering*: Similar to classification is clustering. The major difference being that no groups have been predefined.
- iii. *Prediction*: 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.
- iv. *Forecasting*: 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. 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.

III. DATA MINING PROCESS BASED ON NEURAL NETWORK

Data mining process can be composed by three main phases:

- A. data preparation,
- B. data mining,
- C. expression and interpretation of the results,

Data mining process is the reiteration of the three phases. The details are shown in Fig. 4.

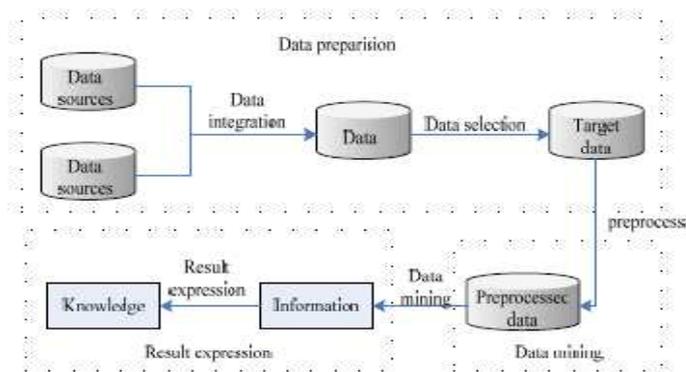


Figure 4. General data mining process.

The data mining based on neural network is composed by data preparation, rules extracting and rules assessment three phases, as shown in Fig. 5.

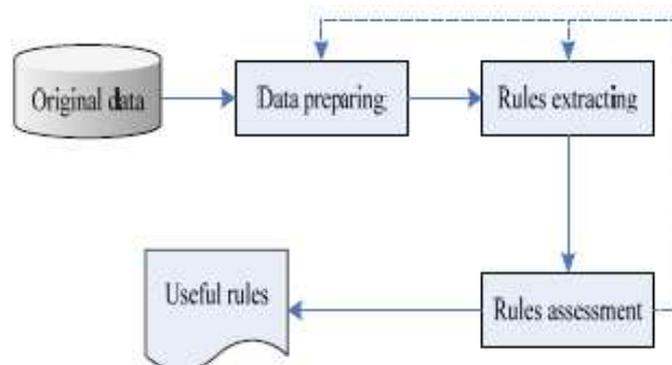


Figure 5. Data mining process based on neural network.

A. Data Preparation

Data preparation is to define and process the mining data to make it fit specific data mining method. Data preparation is the first important step in the data mining and plays a decisive role in the entire data mining process. It mainly includes the following four processes:

- a) *Data cleaning*: Data cleansing is to fill the vacancy value of the data, eliminate the noise data and correct the inconsistencies data in the data.
- b) *Data option*: Data option is to select the data arrange and row used in this mining.
- c) *Data preprocessing*: Data preprocessing is to enhanced process the clean data which has been selected.
- d) *Data expression*

Data expression is to transform the data after preprocessing into the form which can be accepted by the data mining algorithm based on neural network. The data mining based on neural network can only handle numerical data, so it is need to transform the sign data into numerical data. The simplest method is to establish a table with one-to-one correspondence between the sign data and the numerical data.

The other more complex approach is to adopt appropriate Hash function to generate a unique numerical data according to given string. Although there are many data types in relational database, but they all basically can be simply come down to sign data, discrete numerical data and serial numerical data three logical data types. Fig. 6 gives the conversion of the three data types. The symbol “Apple” in the figure can be transformed into the corresponding discrete numerical data by using symbol table or Hash function. Then, the discrete numerical data can be quantified into continuous numerical data and can also be encoded into coding data.

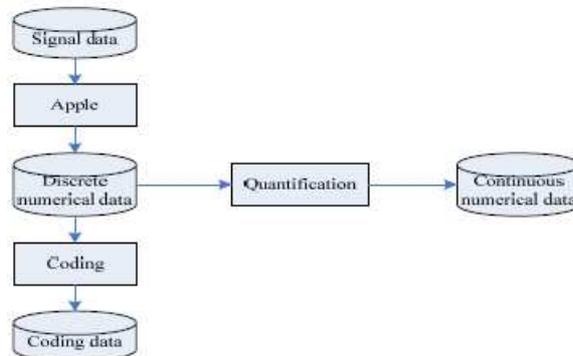


Fig. 6 Data expression and conversion in data mining based on neural network

B. Rules Extracting

There are many methods to extract rules, in which the most commonly used methods are LRE method, black-box method, the method of extracting fuzzy rules, the method of extracting rules from recursive network, the algorithm of binary input and output rules extracting (BIO-RE), partial rules extracting algorithm (Partial-RE) and full rules extracting algorithm (Full-RE).

C. Rules Assessment

Although the objective of rules assessment depends on each specific application, but, in general terms, the rules can be assessed in accordance with the following objectives.

- 1) Find the optimal sequence of extracting rules, making it obtains the best results in the given data set;
- 2) Test the accuracy of the rules extracted;
- 3) Detect how much knowledge in the neural network has not been extracted;
- 4) Detect the inconsistency between the extracted rules and the trained neural network.

IV. CONCLUSION

At present, neural network 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. Compared to statistical methods, NN are useful especially when there is no a priori knowledge about the analyzed data. They offer a powerful and distributed computing architecture, with significant learning abilities and they are able to represent highly nonlinear and multivariable relationships.

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.

V. 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] Haykin, S., *Neural Networks*, Prentice Hall International Inc., 1999.
- [5] Khajanchi, Amit, *Artificial Neural Networks: The next intelligence*.
- [6] Zurada J.M., "An introduction to artificial neural networks systems", St. Paul: West Publishing (1992).
- [7] Y. Bengio, J. M. Buhmann, M. Embrechts, and J.M. Zurada. Introduction to the special issue on neural networks for data mining and knowledge discovery. *IEEE Trans. Neural Networks*.
- [8] M. W. Craven and J. W. Shavlik. Using neural networks for data mining. *Future Generation Computer Systems*, 13:211–229, 1997.