Data Mining Using Neural Network

Latika Bokadiya
Assist. Prof. (Computer Deptt)
S.I.T.E, Nathdwara(Raj.)
latikabokadiya@gmail.com

Anand kumar
Assist. Prof. (Computer Deptt)
S.IT.E, Nathdwara(Raj.)
a.kumar_site@yahoo.com

Sachin Chauhan
Scholar B.TECH(CS)
S.I.T.E Nathdwara(Raj.)
slc1312@gmail.com

Abstract:-*The benefit of neural networks in the data mining has become wider. 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 research 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 and analyzed.*

Keywords: - Artificial Neural Network (ANN), neural network topology, back propagation algorithm, Data mining; neural networks, data mining processes, implementation.

1. INTRODUCTION:-

Due the continuous development of database technology and the excessive applications of database management system, the data volume stored in database increases rapidly and in the huge amounts of data much important information is hidden. If the information can be extracted from the database they will create a lot of potential profit for the companies, and the technology of mining information from the extensive database is known as data mining.

Data mining tools can prognosis the future trends and activities to support the accord of people. For example, analyzing the complete database system of the organization the data mining tools can answer the difficulties such as "Which user is most likely to respond to the e-mail marketing activities of our organization, why", and other similar difficulties. Some tools of data mining can also resolve some traditional problems which consumed too much time, because that they can expeditiously browse the entire database and find some useful information experts unnoticed.

Neural network is also called parallel processing network which generated with simulating the image perceptive thinking of human, on the basis of the research of biological neural network, according to the features of biological neurons and neural network and by sampling, summarizing and refining. It uses the idea of non-linear aligning, the method of parallel processing and the architecture of the neural network itself to represent the associated knowledge of input and output. Initially, the use of the neural network in data mining was not optimistic, and the main reasons are that the neural network has the defects of complex architecture, poor interpretability and long training time. But its advantages such as high affordability to the noise data and low error rate, the continuously enhancement and optimization of different network training algorithms, mainly the continuously advancing and improvement of different network pruning algorithms and rules extracting algorithm, make the application of the neural network in the data mining increasingly supported by the overwhelming majority of users. In this paper the data mining based on the neural network is researched in detail.

2. NEURAL NETWORK METHOD:

There are seven common approaches and techniques of data mining which are the methods of statistical analysis, rough view, covering positive and rejecting inverse cases, fuzzy method, formula found, as well as visualization technology. Now here, we focus on neural network method.

Neural network method is used for clustering, classification, feature mining, prediction and pattern recognition. It simulate 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 working out data mining, the neural network method gradually calculates (including repeated iteration or cumulative computation) the weights the neural network connected. The neural network model can be broadly divided into the following three types:

(i) Feed-forward networks: it regards the concept back-propagation structure and the function network as representatives, and mainly used in the areas such as prediction and pattern recognition.

(ii) Feedback network: it regards discrete model and continuous model as representatives, and specially used for associative memory and optimization estimation.

(iii) Self-organization networks: it respects adaptive. At present, the neural network most commonly used in data

mining is BP network. Of course, ANN is the developing science, and some theories have not really taken shape, such as the problems of stability, convergence, local minimum and parameters readjustment. For the BP network the frequent problems it encountered are that the working is slow, may fall into local minimum and it is very difficult to determine training parameters. Aiming at these problems some people adopted the method of combining ANN and genetic gene algorithms and achieved better results.

3. DATA MINING PROCESS DEPEND ON NEURAL NETWORK:-

Data mining process can be done by three main phases: data preparation, data mining, expression and interpretation of the results, data mining process is the reiteration of the three phases. The details are shown in fig 1.

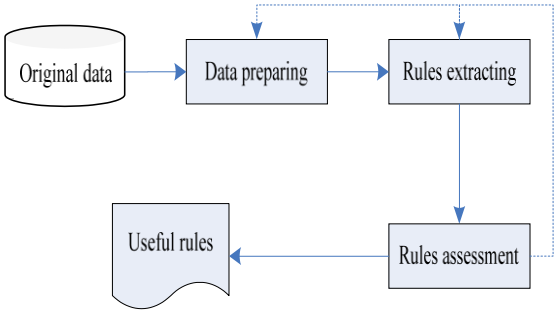

Fig.1 General Data mining Process

The data mining based on neural network is composed by data preparation, rules extracting and rules assessment three phases as shown in fig. 2.

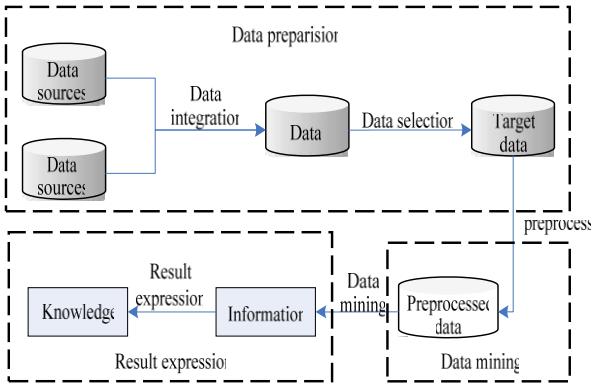

Fig.2 Data mining process based on neural network

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

(i) Data cleaning: Data cleansing is to fill the missing and vacancy value of the data, eliminate the noise and undesired data and correct the inconsistencies data in the data.

(ii) Data option: Data option is to select the data to arrange and row used in this mining.

(iii) Data preprocessing: Data preprocessing is to enhanced method the clean data which has been selected.

(iv) Data expression: Data expression is to convert the data after preprocessing into the form which can be used and accepted by the data mining algorithm based on neural network. The data mining (DM) based on neural network can only handle numerical data, so it is necessary to transform the sign data into numeral data. The easiest method is to demonstrate a table with one-to-one correspondence between the sign data and the numerical data. The some other more difficult approach is to adopt appropriate Hash function to generate a most unique numerical data according to given string.

Although there are many data types in relational database, but they all fundamentally can be simply come down to sign data, discrete numerical data and serial numeral data three logical data types. Fig. 3 gives the conversion of the three data types.

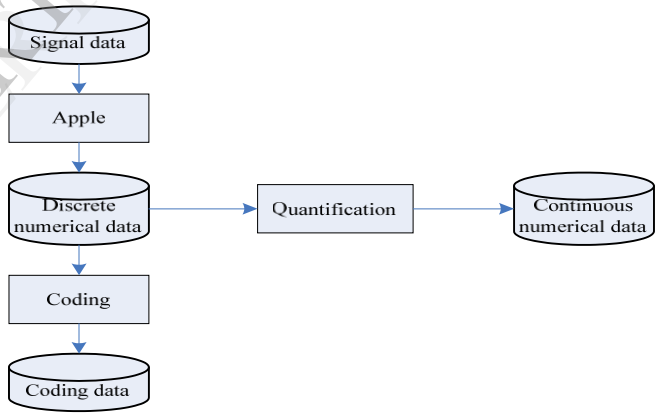

Fig. 3 Data expression and conversion in data mining based on neural network

The symbol “Apple” in the figure can be transformed into the corresponding discrete numeral data by using symbol table or Hash function. Then, the discrete numeral data can be quantified into continuous numeral data and can also been coded into coding data.

B. Rules Extracting

There are many techniques 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 appraisal depends on each specific application, but, in general terms, the rules can be assessed in accordance with the following objectives.

- Find the optimal sequence of extracting rules, making it obtains the best outputs in the given data set;
- Test the precision of the rules extracted;
- Detect how much cognition in the neural network has not been extracted;
- Detect the incompatibility between extracted rules and the trained neural network.

4. TYPES BASED DATA MINING ON NEURAL NETWORK

The types of knowledge mining supported neural network square measure tons of, however there square measure solely 2 varieties most used that square measure the info mining supported the self-organization neural network and on the fuzzy neural network.

A. information Mining primarily based on Neural-Network Self-organization method. It is not a method of learning while not lecturers. Through the study, the vital characteristics or some inherent data during a cluster of knowledge, like the characteristics of the distribution or bunch according to bound feature. Scholars T. Kohonen of Finland considers that the neighboring modules within the neural network square measure almost like the brain neurons and play totally different rules, through interaction they will be adaptively developed to be special sight or to detect totally different signal. as a result of the brain neurons {in totally different| in several |in numerous} brain area elements play different rules, in order that they square measure sensitive to totally different input modes. T Kohonen conjointly planned a sort of learning mode that makes the signaling be mapped to the low-dimensional area, and maintain that the signaling with same characteristics may be similar to regional region in area, that is that the supposed self-organization feature map (S0FM).

B. data processing supported Fuzzy Neural Network:

Although neural network has robust functions of learning, classification, association and memory, however within the use of the neural network for data processing, the best issue is that the output results can't be intuitively well-lighted. Once the introduction of the fuzzy process performs into the neural network, it can't solely increase its output expression capability however additionally the system becomes a lot of stable. The fuzzy neural networks often employed in data processing ar fuzzy perception model, fuzzy BP network, fuzzy agglomeration Kohonen network, fuzzy abstract thought network and fuzzy ART model. Within which the fuzzy BP network is developed from the ancient BP network. within the ancient BP network, if the samples belonged to the primary k class, then except the output worth of the primary k output node is one, the output worth of alternative output nodes all is 0, that is, the output worth of the standard BP network solely can be zero or one, is not ambiguous. However, in fuzzy BP networks, the expected output worth of the samples is replaced by the expected membership of the samples like varied varieties. Once coaching the samples and their expected membership like varied varieties in learning stage fuzzy BP network can have the flexibility to replicate the affiliation relation between the input and output in coaching set, and may provide the membership of the popularity pattern in data processing. Fuzzy agglomeration Kohonen networks achieved fuzzy not solely in output expression, however additionally introduced the sample membership into the change rules of the weight constant, that makes the change rules of the burden constant has additionally complete the fuzzy.

Evaluating whether or not has knowledge of data and information. Mining implementation algorithmic program is okay the subsequent indicators and characteristics are used:

(i) Whether or not high-quality modeling beneath the circumstances of noise and data half-baked;

(ii) The model should be understood by users will and may and might be used for decision-making; **(iii)** the model can receive space knowledge (rules enter and extraction) to boost the modeling quality. Existing neural network has high exactness within the quality of modeling however low within the latter 2 indicators. Neural network truly is seen as a recorder for users, the appliance restrictions makes the classification and prediction method cannot be understood by users and directly used for decision-making. For knowledge mining, it not enough to rely on the neural network model providing results as a result of that before vital decision-making users have to be compelled to perceive the explanation and justification for the decision-making. Therefore, within the AN{data mining|data methoding} data base ought to be established in order to accede domain data and also the data ANN learning to the system within the data processing process. that's to mention, within the ANN data processing, it's necessary to use data methodology to extract data from the info mining method and understand the conjugation of the data process and neural network. Additionally, within the system a good call and clarification mechanism ought to even be thought-about to be established to boost the validity and usefulness of the ANN data processing technology.

C. Input/output Interface

Considering that the strategy of exploitation neural network tools or neural network code package to get knowledge is laggard, then a decent interface with relative information, multi-dimensional information and knowledge warehouse ought to be established to fulfill the requirements of knowledge mining.

V. CONCLUSION

At present, data processing may be a new and vital space of analysis, and neural network itself is incredibly appropriate for determination the issues of knowledge mining as a result of its characteristics of fine lustiness, self-organizing adaptive , data processing, distributed storage and high degree of fault tolerance. the mix of knowledge mining methodology and neural network model will greatly improve the potency of knowledge mining strategies, and it's been wide used. It conjointly can receive a lot of and a lot of attention.

REFERENCES:

[1] S Lawrence, C Lee Giles. Accessibility of Information on the Web [J].
[2] Guan Li Liang, Hongjun, data warehouse and Data-Mining,Microcomputer-Applications, 1999, 15(9): 17-20.
[3] Adriaans P, Zantinge D. Data mining [M]. Addision_Wesley Longman, 1996.
[4] Chen Rong, BP arithmetic and its structure optimization tactics,Journal of Autoimmunization. 1997, 23(1), 43-49.
[5] G.Towell, J.W Shavlik. The extraction of refined rules from knowledge-based neural networks [J]. Machine Learning, 1993(13): 71-101.
[6] Yang Kun, Liu Dayou. Agents: properties and classifications. Computer Science [J]. 1999, 26(9): 30-34.
[7] H Lu, R Setiono, H Liu. Effective Data Mining Using Neural Network. IEEE Transactions on Knowledge and Data Engineering, , 1996, 8(6):957-961.
[8] David Hand, Principles of Data Mining [M]. Massachusetts institute of technology,2001.
[9] Feng Jiansheng. KDD and its applications, BaoGang techniques, 1999(3): 27-3.
[10] Wooldrldge M J. Agent-Based software engineering Agent-Based software engineering. IEEE Transactions on Software Engineering [J]. 1999,144 (1):26-27.

.