Application of Back Propagation Neural Network on Requirement Data Mining

1Rekhanjali Sahoo, 2Brojo Kishore Mishra

1Department of CSE, Gurukul Institute of Technology, Bhubaneswar
2Department of IT, C. V. Raman College of Engineering (Autonomous), Bhubaneswar
Email: 1rekhanjalisahoo23@gmail.com, 2brojokishoremishra@gmail.com

Abstract: Data mining is a knowledge discovery approach. It deals with large databases that impose on clustering analysis additional severe computational requirements. It is also the process of extracting knowledge or predicting previously unknown and useful trends from large quantities of data. The knowledge of extracting data by using multidisciplinary fields such as statistics, mode identifies artificial intelligence, machine learning, database and so on.

In this paper, the back propagation (BP) neural network method is used as the technique of data mining to analyze the effects of structural technologic parameters on efficiency in resume filtering. The main aim is to choose effective candidates for better development of the organization by applying back propagation algorithm. Also here we categorize the students into grade order in all their education studies and it helps in interview situation. This study includes the variables such as age, gender, name, lower class grade, higher class grade, degree proficiency and extra knowledge or skill, etc. It helps to categorize students in rank order to arrange for the recruitment process. Due to this, it reduces the short listings. Here, clustering, association rules, classification and outlier detection has been used to evaluate the students performance.

Index Terms— Data mining, Clustering, Classification, Association rule, Outlier detection, Preprocessing, Artificial Neural Network, Back propagation algorithm.

I. INTRODUCTION

Data mining deals with large databases that impose on clustering analysis. Data mining concepts and methods can be applied in various fields. In Educational data mining it is a new technique of data mining that can be applied on the data related to the field of education. It uses different kinds of knowledge using association rules, classification and clustering, neural networks. In the face of huge amounts of data, the first task is to sort them out. So Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups. Here we use the concept of neural network. A Neural Network is a powerful data-modeling tool that is able to capture and represent complex input/output relationships. It performs similar tasks like human brain. The artificial neural network (ANN) is one of the most efficient techniques of data mining. ANN has the ability to mapping high nonlinear system, asociable memory and abstractly generalization. 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. Back propagation is a form of supervised learning for multi-layer nets, also known as the generalized delta rule. Error data at the output layer is “back propagated” to earlier ones, allowing incoming weights to these layers to be updated. It is most often used as training algorithm in current neural network applications. The back propagation algorithm is an involved mathematical tool. Artificial Neural change its weights according to some learning rule. Supervised learning or Associative learning which incorporates an external teacher, so that each output unit is told what its desired response to input signals ought to be. The type of learning is determined by the manner in which the parameter changes take place. All learning methods used for neural networks can be classified into two major categories: SUPERVISED LEARNING which incorporates an external teacher, so that each output unit is told what its desired response to input signals ought to be. UNSUPERVISED LEARNING uses no external teacher and is based upon only local information. It is also referred to as self-organization. In this paper, results of the experiments conducted to cluster the data with the K means clustering and classification using back propagation algorithm have been analyzed. A number of other clustering algorithms have been developed. Some deal with the specific application requirements. Constraint-based clustering belongs to this category One of the mechanisms used by the IT industries is to conduct test and group discussion, marks obtained in the semester exams and etc. On the other hand institutes has different criteria such as Marks obtained, experience, higher qualification, demo, communication skill etc according to their or universities rules. It has been observed that lot of resumes are collected only for 1 or 2 posts. The time taken to conduct interview of all the candidates consumes more time. And lot of efforts is put to analyze the profile of the all applications to determine.
II. ARCHITECTURE AND WORK

The design of the system requires the complete understanding of the problem domain. The data sets and the input attributes are determined through knowledge of an engineering college. The task and purpose should be confirmed at the beginning of data-mining stage. What algorithm would be used in mining is determined according to the characteristics of the task type. The back propagation algorithm is a widely used method in neural networks.

A. Architecture of Neural Networks

i. NETWORK LAYERS

1. The commonest type of artificial neural network consists of three groups, or layers, of units: a layer of "input" units is connected to a layer of "hidden" units, which is connected to a layer of "output" units.
2. The activity of the input units represents the raw information that is fed into the network.
3. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units.
4. The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

ii. FEED-FORWARD NETWORKS

1. Feed-forward ANN's allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed forward ANN's tend to be straightforward networks that associate inputs with outputs.
2. They are extensively used in pattern recognition. This type of organization is also referred to as bottom-up or top-down.

iii. FEEDBACK NETWORKS

1. Feedback networks can have signals traveling in both directions by introducing loops in the network.
2. Feedback networks are very powerful and can get extremely complicate.

The complete back propagation algorithm can now formulate and prove by induction that it works in arbitrary feed-forward networks with differentiable activation functions at the nodes. We assume that we are dealing with a network with a single input and a single output unit.

III. BACK PROPAGATION ALGORITHM.

The back propagation algorithm was developed by Paul Werbos in 1974 and rediscovered independently by Rumelhart and Parker. Since its rediscovery, the back propagation algorithm has been widely used as a learning algorithm in feed forward multilayer neural networks.

1. Consider a network with a single real input x and network function F. The derivative \( F'(x) \) is computed in two phases:

2. Feed-forward: the input x is fed into the network.

3. The primitive functions at the nodes and their derivatives are evaluated at each node.

4. The derivatives are stored. Back propagation: the constant 1 is fed into the output unit and the network is run backwards. Incoming information to a node is added and the result is multiplied by the value stored in the left part of the unit. The result is transmitted to the left of the unit.

5. The result collected at the input unit is the derivative of the network function with respect to x.

The aim is to determine a set of weights and threshold values that satisfy the precision Requirement.
Another system was the ADALINE (Adaptive Linear Element) which was developed in 1960 by Widrow and Hoff (of Stanford University). The ADALINE was an analogue electronic device made from simple components. The method used for learning was different to that of the Perceptron.

Fukushima (F. Kunihiko) developed a step wise trained multilayered neural network for interpretation of handwritten characters. The original network was published in 1975 and was called the Cognitron.

Mohammed M. AbuTair, Alaa M. EL-Halees [2], had a survey on educational data mining [1993-2007] they collected graduate students information and applied data mining techniques to discover knowledge. Using discovered association rules, they sorted the rules using lift metric. Then they used two way classification methods which are rule induction and naive Bayesian classifier to predict the grade of graduate students.

Romero and Ventura[5], did on a survey on educational data mining between 1995 and 2005. They concluded that educational data mining is a promising area of research and it has specific requirements not presented in other domains.

IV. RESULT ANALYSIS

Association rule generation is usually split up into two separate steps: First, minimum support is applied to find all frequent item sets in a database. Second, these frequent item sets and the minimum confidence constraint are used to form rules. Finding all frequent item sets in a database is difficult since it involves searching all possible item sets (item combinations). The set of possible item sets is an I and has size 2^n-1(excluding the empty set which is not a valid item set). Figure 2 depicts a sample of association rules discovered from data students with average grade, with their support, confidence [2].

\[
\text{[Lower class grade}=\text{Poor}, \quad \text{Higher class grade}=\text{Good}] \rightarrow \quad \text{[Grade}=\text{Average}] \\
\text{Support: 0.19, \quad Confidence: 0.757)
\]

\[
\text{[Lower class grade}=\text{Good}, \quad \text{Higher class grade}=\text{Poor}] \rightarrow \quad \text{[Grade}=\text{Average}] \\
\]

A. SYNAPSE

INPUTS AND OUTPUTS FOR THE ENTITY SYNAPSE

i. INPUTS

\[
w_{11}, w_{12}, w_{13}, w_{14}, w_{21}, w_{22}, w_{23}, w_{24}, w_{31}, w_{32}, w_{33}, w_{34}, w_{41}, w_{42}, w_{43}, w_{44} \text{ are the input weights,}
\]

\[
i_1, i_2, i_3, i_4 \text{ are the input signals,} \quad \text{start, m are the input}
\]

signals introduced to take care of timing constraints during port mapping.

ii. OUTPUTS

\[
s_1, s_2, s_3, s_4 \text{ are the sum of weighted inputs,}
\]

\[
m_{11}, m_{12}, m_{13}, m_{14}, m_{21}, m_{22}, m_{23}, m_{24}, m_{31}, m_{32}, m_{33}, m_{34}, m_{41}, m_{42}, m_{43}, m_{44} \text{ are the intermediate signals obtained by multiplying inputs and the corresponding}
\]

Weights done are the output signals introduced to take care of timing constraints during port mapping.

V. CONCLUSION

In this paper we present a neural network based approach to mining classification rules from given databases. The approach consists of different phases:

1) Constructing and training a network to correctly classify tuples in the given training data set to required accuracy,

2) Extracting knowledge through the network using BPNN

A set of experiments was conducted to test the proposed approach is using a well defined set of data mining problems. The results indicate that, using the proposed approach, high quality or useful data can be discovered from the given data sets. In future we can apply this technique to select deserving candidates for an organization. This technique is also applicable on feedback system in any organization. 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 we have to find the right candidates for organization. In future we can apply this technique to select deserving candidates for an organization. This technique is also applicable on feedback system in any organization. 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.

In this paper data mining techniques are efficiently used to categorize the level of students. One of the data mining techniques that is classification, accurately classifies the
data for categorizing student based on the levels. As one important function of data mining, clustering analysis either as a separate tool to discover data sources distribution of information, as well as other data mining algorithm as a preprocessing step, the cluster analysis has been into the field of data mining is an important research topic. Clustering is used to the group the students according to their grade and proficiency. This goes a long way to help how define the recruitment process in a easier manner

REFERENCES


