

DATA MINING THROUGH NEURAL NETWORKS USING RECURRENT NETWORK

Gaurab Tewary

MCA , GGSIPU, New Delhi
tewary.gaurab@gmail.com

ABSTRACT

With the development of database, the data volume stored in database increases rapidly and in the large amounts of data much important information is hidden. If the information can be extracted from the database they will create a lot of profit for the organization. The question they are asking is how to extract this value. The answer is data mining. There are many technologies available to data mining practitioners, including Artificial Neural Networks, Genetics, Fuzzy logic and Decision Trees. Many practitioners are wary of Neural Networks due to their black box nature, even though they have proven themselves in many situations. This paper is an overview of artificial neural networks and questions their position as a preferred tool by data mining practitioners.

KEYWORDS

ANN- Artificial Neural Networks, ESRNN- Extraction of Symbolic Rules from ANN's, data mining, symbolic rules

1. 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. Following example of a financial institution failing to utilize their data-warehouse. Income is a very important socio-economic indicator. If a bank knows a person's income, they can offer a higher credit card limit or determine if they are likely to want information on a home loan or managed investments. Even though this financial institution had the ability to determine a customer's income in two ways, from their credit card application, or through regular direct deposits into their bank account, they did not extract and utilize this information [1,2]

An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. ANN is an adaptive system that changes its structure based on external or internal information that flows through the

network during the learning phase. They are used to model complex relationships between inputs and outputs or to find patterns in data. Example Facial or Handwriting or Voice Recognition[3].

The Recurrent Structure is also known as Auto associative or Feedback Network, they contain feedback connections Contrary to feed forward neural network. It regards Competitive model etc., and mainly used for associative memory and optimization calculation[5,6].

In this paper we discuss a data mining scheme, referred to as ESRNN (Extraction of Symbolic Rules from ANNs) to extract symbolic rules from trained ANNs. A three-phase training algorithm. In the first and second phases, appropriate network architecture is determined using weight freezing based constructive and pruning algorithms. In the third phase, symbolic rules are extracted using the frequently occurred pattern based rule extraction algorithm by examining the activation values of the hidden nodes[10].

2. INTRODUCTION OF DATA MINING

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. Example of a financial institution failing to utilize their data-warehouse is in cross-selling insurance products (e.g. home, life and motor vehicle insurance). By using transaction information they may have the ability to determine if a customer is making payments to another insurance broker. This would enable the institution to select prospects for their insurance products.[1,2]

2.1 Need of Data Mining

Finding information hidden in data is as theoretically difficult as it is practically important. With the objective of discovering unknown patterns from data, Companies have been collecting data for decades, building massive data warehouses in which to store it. Even though this data is available, very few companies have been able to realize the actual value stored in it. The question these companies are asking is how to extract this value. The answer is Data mining[1,2]

2.2 Techniques/Functionalities of Data Mining

There are two fundamental goals of data mining: prediction and description. Prediction makes use of existing variables in the database in order to predict unknown or future values of interest, and description focuses on finding properties that describe the existing data.[3]. There are several data mining techniques fulfilling these objectives. Some of these are associations, classifications, sequential patterns and clustering. Another approach of the study of data mining techniques is to classify the techniques as: user-guided or verification-driven data mining and, discovery-driven or automatic discovery of rules.

A. Association Rules:

An association rule is an expression of the form $X \Rightarrow Y$, where X and Y are the sets of items. The meaning of such a rule is that the transaction of the database, which contains X tends to contain Y . Given a database, the goal is to discover all the rules that have the support and confidence greater than or equal to the minimum support and confidence, respectively.

Support means how often X and Y occur together as a percentage of the total transactions. Confidence measures how much a particular item is dependent on another. Patterns with a combination of intermediate values of confidence and support provide the user with interesting and previously unknown information.

B. Classification Rules:

Classification involves finding rules that partition the data into disjoint groups. The input for the classification data set is the training data set, whose class labels are already known. Classification analyses the training data set and constructs a model based on the class label, and aims to assign class label to the future unlabelled records. Since the class field is known, this type of classification is known as supervised learning. There are several classification discovery models. They are: the decision tree, neural networks, genetic algorithms and some statistical models.

C. Clustering :

Clustering is a method of grouping data into different groups, so that the data in each group share similar trends and patterns. The goal of the process is to identify all sets of similar examples in the data, in some optimal fashion. If a measure of similarity is available, then there are a number of techniques for forming clusters. It is an Unsupervised classification.

Heuristic Clustering Algorithm[10]

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar within the same cluster and are dissimilar to the objects in other clusters. A cluster of a data objects can be treated collectively as one group in many applications. There exist a large number of clustering algorithms, such as, k-means, k-memoids. The choice of clustering algorithm depends both on the type of data available and on the particular purpose and applications.

After applying pruning algorithm in ESRNN, the ANN architecture produced by the weight freezing based constructive algorithm contains only important nodes and connections. Therefore, rules are not readily extractable because the hidden node activation values are continuous. The separation of these values paves the way for rule extraction. It is found that some hidden nodes of an ANN maintain almost constant output while other nodes change continuously during the whole training process. Figure shows output of three hidden nodes where a hidden node maintains almost constant output value after some training epochs but output value of other nodes are changing continually. In ESRNN, no clustering algorithm is used when hidden nodes maintain almost constant output value. If the outputs of hidden nodes do not maintain constant value, a heuristic clustering algorithm is used.

The aim of the clustering algorithm is to separate the output values of the hidden nodes. Consider that the number of hidden nodes in the pruned network is H . Clustering the activation values of the hidden node is accomplished by a simple greedy algorithm that can be summarized as follows:

1. Find the smallest positive integer d such that if all the network activation values are rounded to d decimal places, the network still retains its accuracy rate

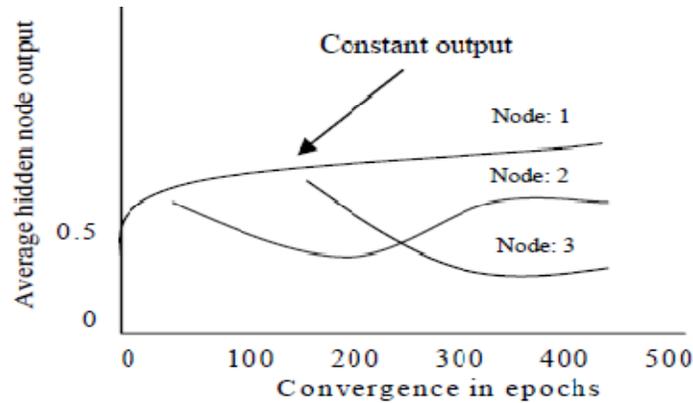


Figure 1. Output of the hidden nodes.

2. Represent each activation value α by the integer closest to $\alpha \times 10d$. Let $H_i = \langle h_{i,1}, h_{i,2}, \dots, h_{i,k} \rangle$ be the k -dimensional vector of these representations at hidden node i for patterns x_1, x_2, \dots, x_k and let $H = (H_1, H_2, \dots, H_H)$ be the $k \times H$ matrix of the hidden representations of patterns at all H hidden nodes.
3. Let P be a permutation of the set $\{1, 2, \dots, H\}$ and set $m = 1$.
4. Set $i = P(m)$.
5. Sort the values of the i th column (H_i) of matrix H in increasing order.
6. Find a pair of distinct adjacent values $h_{i,j}$ and $h_{i,j+1}$ in H_i such that if $h_{i,j+1}$ is replaced by $h_{i,j}$ no conflicting data will be generated.
7. If such a pair of values exists, replace all occurrences of $i, j, h + in$ in H_i by i, j, h and repeat Step 6. Otherwise, set $m = m + 1$. If $m \leq H$, go to Step 4, else stop.

The activation value of an input pattern at hidden node m is computed as the hyperbolic tangent function, it will have a value in the range of $[-1, 1]$. Steps 1 and 2 of the clustering algorithm find integer representations of all hidden node activation values. A small value for d in step 1 indicates that relatively few distinct values for the activation values are sufficient for the network to maintain its accuracy.

The array P contains the sequence in which the hidden nodes of the network are to be considered. Different ordering sequences usually result in different clusters of activation values. Once a hidden node is selected for clustering, the separated activation values are sorted in step 5 such that the activation values are in increasing order. The values are clustered based on their distance. We implemented step 6 of the algorithm by first finding a pair of adjacent distinct values with the shortest distance. If these two values can be merged without introducing conflicting data, they will be merged. Otherwise, a pair with the second shortest distance will be considered. This process is repeated until there are no more pairs of values that can be merged. The next hidden node as determined by the array P will then be considered.

2.3 Challenges of Data Mining

- 1) The whole Data Mining process consumes a large amount of time.
- 2) Data Mining is Expensive. .
- 3) Classification in Data Mining.
- 4) The whole Data Mining process depends on a proper valid input, without a proper input Data Mining process cannot produce a proper valid output.

3. INTRODUCTION OF NEURAL NETWORKS

An Artificial Neuron is basically an engineering approach of biological neuron. It has device with many inputs and one output. ANN is consist of large number of simple processing elements that are interconnected with each other and layered also. In human body work is done with the help of neural network. Neural Network is just a web of inter connected neurons which are millions and millions in number. With the help of this interconnected neurons all the parallel processing is done in human body and the human body is the best example of Parallel Processing. Example Facial or Handwriting or Voice Recognition[6] A neuron is a special biological cell that process information from one neuron to another neuron with the help of some electrical and chemical change. It is composed of a cell body or soma and two types of out reaching tree like branches: the axon and the dendrites. The cell body has a nucleus that contains information about hereditary traits and plasma that holds the molecular equipments or producing material needed by the neurons. The whole process of receiving and sending signals is done in particular manner like a neuron receive signals from other neuron through dendrites. The Neuron send signals at spikes of electrical activity through a long thin stand known as an axon and an axon splits this signals through synapse and send it to the other neurons.[6]

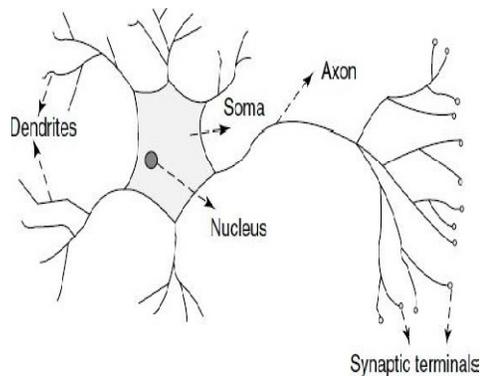


Fig 2 Human Neurons

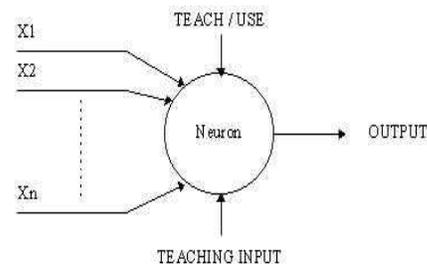


Fig 3 Artificial Neuron

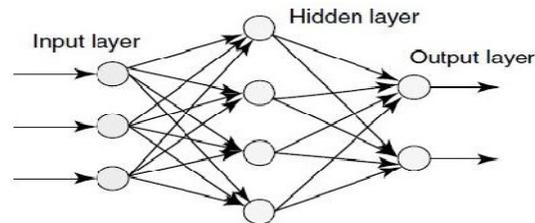


Fig 4 Multilayered ANN

3.1 Characteristics of Neural Networks

The Characteristics are basically those which should be present in intelligent System like robots and other Artificial Intelligence Based Applications. There are six characteristics of Artificial Neural Network which are basic and important for this technology which are showed with the help of diagram:-

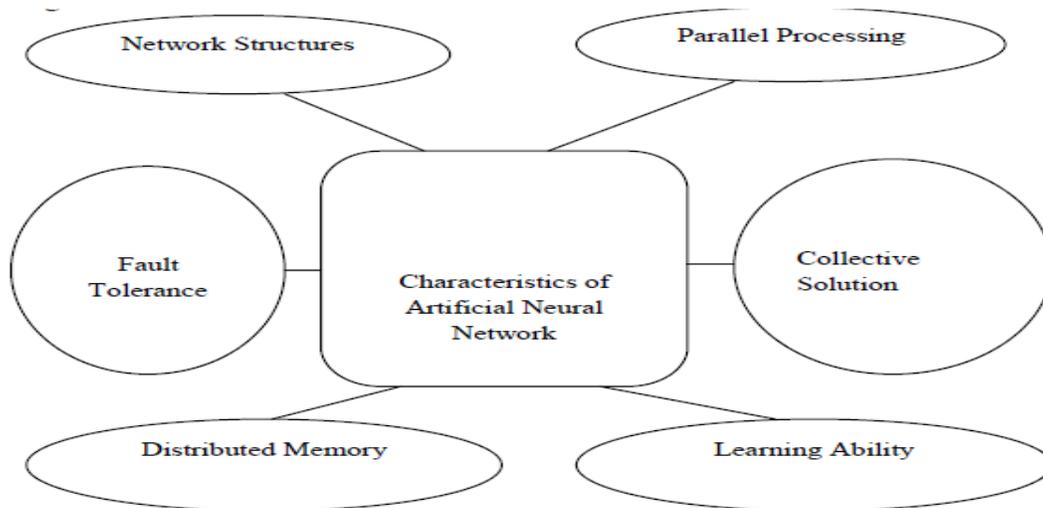


Fig 5 Characteristics

A. The Network Structure:-

There are basically two types of structures recurrent and non recurrent structure. The Recurrent Structure is also known as Auto associative or Feedback Network and the Non Recurrent Structure is also known as Associative or Feed forward Network. In Feed forward Network, the signal travel in one way only but in Feedback Network, the signal travel in both the directions by introducing loops in the network. The Recurrent Structure is also known as Auto associative or Feedback Network, they contain feedback connections Contrary to feed forward neural network. It regards Competitive model etc., and mainly used for associative memory and optimization calculation[5,6].

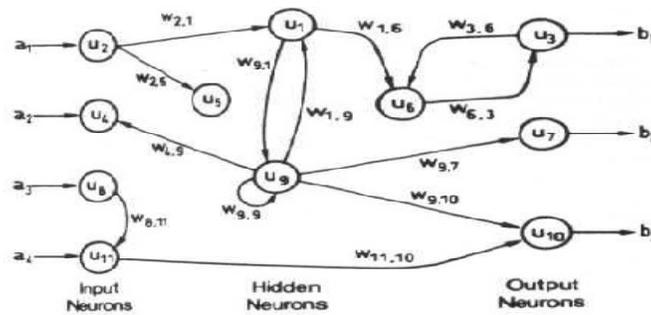


Fig 6 Feed Back Network

Unsupervised Training.[7]

In unsupervised training, the network is provided with inputs but not with desired outputs. The system itself must then decide what features it will use to group the input data. This is often referred to as self-organization or adaption. It is of two types Hebbian & Competitive.

Competitive Networks [6,8]

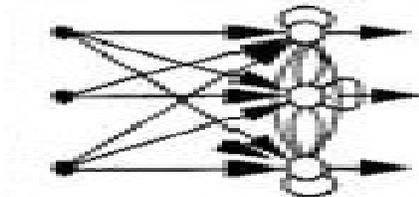


Fig 7 competitive networks

In it output units compete among themselves for activation. As a result, only one output unit is active at any given time. This phenomenon is known as winner-take-all. It has been found to exist in biological neural network. Competitive learning often clusters or categorizes the input data. Similar patterns are grouped by the network and represented by a single unit. This grouping is done automatically based on data correlations. The simplest competitive learning network consists of a single layer of output units.

Hebbian Rule. [8,9]

The oldest learning rule is Hebb's learning(1949).Hebb based it on the following observation from neurobiological experiments: If neurons on both sides of a synapse are activated synchronously and repeatedly, the synapse's strength is selectively increased.

The Hebbian rule can be described as

$$w_i(\text{new}) = w_i(\text{old}) + x_i o,$$

where o is the desired output for

$$i = 1 \text{ to } n(\text{inputs}).$$

An important property of this rule is that learning is done locally, i.e., the change in synapse weight depends only on the activities of the two neurons connected by it.

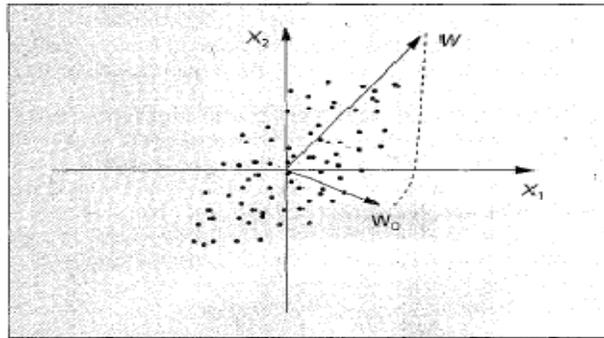


Figure 8. Orientation selectivity of a single neuron trained using the Hebbian rule.

A single neuron trained using the Hebbian rule exhibits an orientation selectivity. Figure 10 demonstrates this property. The weight vector of the neuron is initialized to w_0 , as shown in the figure. As the learning proceeds, the weight vector moves progressively closer to the direction w of maximal variance in the data. In fact, w is the eigenvector of the covariance matrix of the data corresponding to the largest eigen value. Unfortunately, plain Hebbian learning continually strengthens its weights without bound (unless the input data is properly normalized).

B. Parallel Processing Ability:-

Parallel Processing is done by the human body in human neurons are very complex but by applying basic and simple parallel processing techniques we implement it in ANN like Matrix and some matrix calculations.

C. Distributed Memory:-

ANN is very huge system so single place memory or centralized memory cannot fulfill the need of ANN system so in this condition we need to store information in weight matrix which is form of long term memory because information is stored as patterns throughout the network structure.

D. Fault Tolerance Ability:-

ANN is a very complex system so it is necessary that it should be a fault tolerant. Because if any part becomes fail it will not affect the system as much but if the all parts fails at the same time the system will fails completely.

E. Collective Solution:-

ANN is a interconnected system the output of a system is a collective output of various input so the result is summation of all the outputs which comes after processing various inputs.

F. Learning Ability:-

In ANN most of the learning rules are used to develop models of processes, while adopting the network to the changing environment and discovering useful knowledge. These Learning methods are Supervised, Unsupervised and Reinforcement Learning.

4. IMPLEMENTATION OF NEURAL NETWORKS IN DATA MINING

Effective Combination of Neural Network and Data Mining Technology:

The technology almost uses the original ANN software package or transformed from existing ANN development tools, the workflow of data mining should be understood in depth, the data model and application interfaces should be described with standardized form, then the two technologies can be effectively integrated and together complete data mining tasks. Therefore, the approach of organically combining the ANN and data mining technologies should be found to improve and optimize the data mining technology.[4]

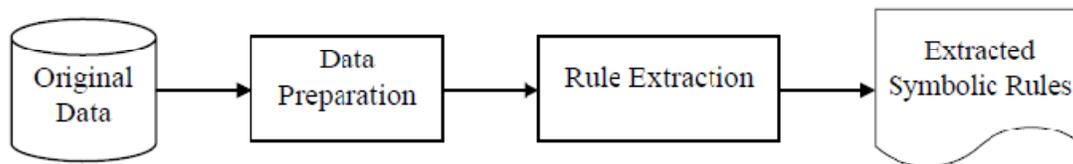


Figure 9. Data mining technique using ANNs.[10,11]

The planned data processing theme consists of two steps: data preparation and rule extraction.

1) Data Preparation

One must prepare quality information by pre-processing the data. The input to the data mining algorithms is assumed to be distributed, containing incorrect values or no missing wherever all options square measure vital. The real-world data could also be noisy, incomplete, and inconsistent, which might disguise helpful patterns. data preparation could be a method of the first information to form it acceptable a particular data mining technique. The data mining using ANNs can only handle numerical data. There are different kinds of attributes that must be representing input and output attributes.

Real-valued attributes square measure sometimes rescaled by some function that maps the value into the range $0 \dots 1$ or $-1 \dots 1$

Integer-valued attributes square measure most often handled as if they were real-valued. If the amount of various values is only small, one among the representations used for ordinal attributes may additionally be applicable.

Ordinal attributes with m different prices are either mapped onto an equidistant scale creating them pseudo-real-valued or are represented by $m - 1$ inputs of that the leftmost k have value 1 to represent the k -th attribute value whereas all others are 0.

5. ANALYSIS OF EXISTING WORK

There are many different approaches for the rule extraction from ANNs that has been developed in the last two decades.[10,11]

Two methods for extracting rules from neural network are described by Towell and Shavlik. The first method is the subset algorithm, which searches for subsets of connections to a node whose summed weight exceeds the bias of that node. The most important downside with subset algorithms is that the price of finding all subsets increases as the size of the ANNs increases. The second method, the M of N algorithm, is an improvement of the set methodology that's designed to expressly seek for M-of-N rules from information based mostly ANNs. Instead of considering an ANN connection, groups of connections are checked for their contribution to the activation of a node, which is done by clustering the ANN connections.

Liu and Tan planned X2R in, an easy and quick algorithmic rule which is be applied to each numeric and discrete data, and generate rules from datasets. It generates good rules within the sense that the error rate of the principles isn't worse than the inconsistency rate found within the original knowledge. The problem of the rules generated by X2R, are order sensitive, *i.e.*, the rules should be fired in sequence.

Afterwards, Setiono presented M of N3, a new method for extracting M-of-N rules from ANNs. The topology of the ANN is the standard three-layered feed forward network. Nodes in the input layer are connected only to the nodes in the hidden layer, while nodes in the hidden layer are also connected to nodes in the output layer. Given a hidden node of a trained ANN with N incoming connections, show how the value of M can be easily computed. In order to facilitate the process of extracting M-of-N rules, the attributes of the dataset have binary values -1 or 1 .

The limitations of the existing rule extraction algorithms are summarized as follows:

- Use predefined and fixed number of hidden nodes that require human experience and prior knowledge of the problem to be solved,
- Clustering algorithms used to separate the output values of hidden nodes are not efficient,
- Computationally expensive,
- Could not produce concise rules, and
- Extracted rules are order sensitive.

6. IMPLEMENTATION OF ESRNN IN NEURAL NETWORKS

Although Artificial Neural Networks (ANNs) have been successfully applied in a wide range of machine learning applications, they are often regarded as “black box”, that means predictions cannot be explained. To enhance the explanation of neural network, a novel algorithm is used known as ESRNN (Extraction of Symbolic Rules from ANNs) to extract symbolic rules from trained ANNs.[10,11]

Extracting symbolic rules from trained ANN is one of the promising areas that are commonly used to explain the functionality of neural network. It is difficult to find the explicit relationship between the input tuples and the output tuples. A number of reasons contribute to the difficulty of extracting rules from a pruned network.

First, even with a pruned network, the links may be still too many to express the relationship between an input tuples and its class label in the form of *if . . . then ...* rules. If a network still has n input links with binary values, there could be as many as 2^n distinct input patterns. The rules could be quite lengthy or complex even for a small n .

Second, a standard ANN is the basis of the proposed ESRNN algorithm. The hyperbolic tangent function, which may take any worth in the interval $[-1, 1]$ is used as the hidden node activation function. Rules are extracted from near optimal neural network by using a new rule extraction algorithm. The aim of ESRNN is to search for simple rules with high predictive accuracy.

The major steps of ESRNN are summarized in Figure:

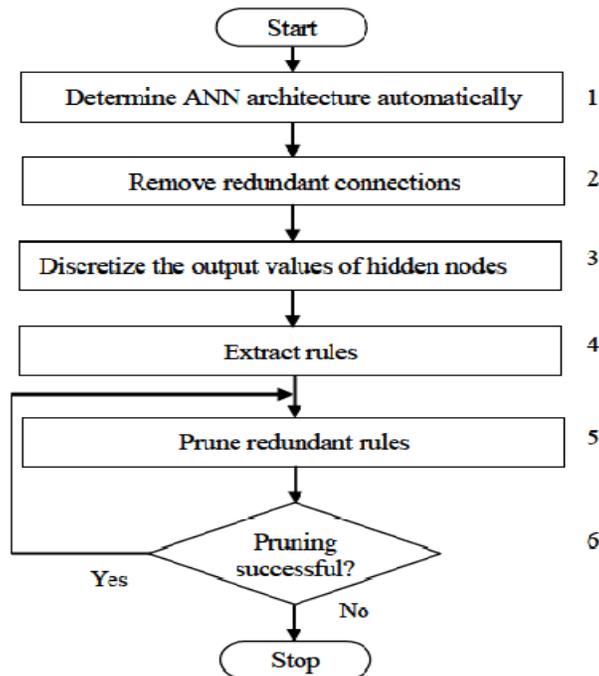


Figure 10. Flow chart of the proposed ESRNN algorithm.

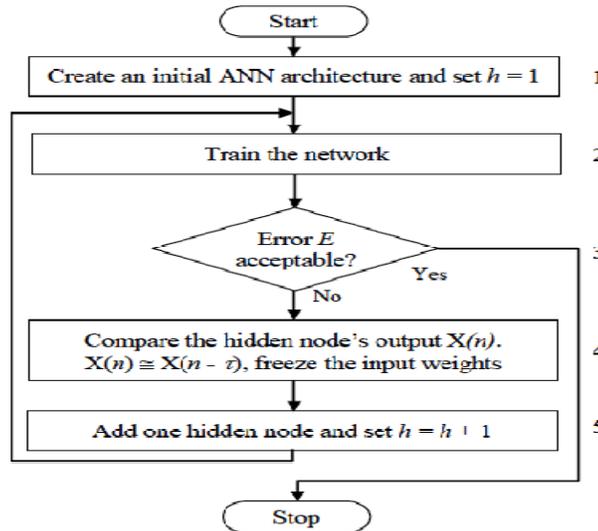
The rules extracted by ESRNN are compact and understandable, and do not involve any weight values. The accuracy of the principles from pruned networks is as high because the accuracy of the original networks. The important features of the ESRNN algorithm are the principles extracted by rule extraction algorithm is recursive in nature and is order insensitive, that is the rules need not to be required to fire sequentially.

6.1 Weight Freezing Based Constructive Algorithm[10,11]

One drawback of the traditional back propagation algorithm is the need to determine the quantity of nodes within the hidden layer prior to training. To beat this issue, several algorithms that construct a network dynamically have been proposed such as DNC, FNNC, CC. However, it is impractical to urge 100% classification accuracy for many of the benchmark classification issues. & higher classification accuracy on the coaching set does not guarantee the higher generalization ability that is classification accuracy on the testing set. The training time is an important issue in designing neural network. One approach for reducing the quantity of weights to be trained is to train few weights rather than all weights during a network and keep remaining weights mounted, commonly referred to as weight freezing.

The thought behind the weight freezing-based constructive algorithm is to freeze input weights of a hidden node once its output does not modification abundant within the consecutive few training epochs. This weight freezing method should be considered as combination of the two extremes: for training all the weights of neural network and for training the weights of only the newly added hidden node of ANNs. In algorithm, it has been proposed that the output of a hidden node can be frozen when its output does not change much in the successive training epochs. The major steps of weight freezing based constructive algorithm are :

Figure 11
Flow chart of the weight freezing based constructive algorithm.



6.2 Pruning Algorithm[10,11]

The pruning algorithm aims at removing redundant links and units without increasing the classification error rate of the network. A small quantity of units and links left in the network after pruning enable us to extract concise and comprehensible rules. Pruning offers an approach for dynamically determinant associate degree acceptable constellation. Pruning techniques begin by training a larger than necessary network and then eliminate weights and nodes that are deemed redundant. The nodes of the hidden layer are determined by weight freezing based constructive

algorithm, the aim of this pruning algorithm used here is to get rid of as several supernumerary nodes and connections as potential. A node is pruned if all the connections to and from the node are pruned. Typically, ways for removing weights from the network involve adding a penalty term to the error function. It is hoped that by add a penalty term to the error function, supernumerary connections can have small weights, and thus pruning will reduce the complexity of the network considerably. The simplest and most commonly used penalty term is the sum of the squared weights. It has been suggested that faster convergence can be achieved by minimizing the cross entropy function instead of squared error function. This pruning algorithm removes the connections of the ANN according to the magnitudes of their weights. As the eventual goal of the ESRNN algorithm is to get a set of simple rules that describe the classification method, it's vital that every one uncalled-for nodes and connections should be removed. In order to get rid of several connections as possible, the weights of the network should be prevented from taking values that are too large. At an equivalent time, weights of irrelevant connections ought to be inspired to converge to zero. The penalty function is appropriate for these purposes.

The steps of the pruning algorithm are explained as follows:

- **Step 1** Train the network to meet a Pre-specified accuracy level with the condition satisfied by all correctly classified input patterns.

$$\max_p |e_{pi}| = \max_p |S_{pi} - t_{pi}| \leq \eta_1, p = 1, 2, \dots, C. \quad (1)$$

Let n_1 and n_2 be positive scalars such that $(n_1 + n_2) < 0.5$ (n_1 is the error tolerance, n_2 is a threshold that determines if a weight can be removed), where $n_1 \in [0, 0.5)$. Let (w, v) be the weights of this network.

- **Step 2** Remove connection between input nodes and hidden nodes, and also remove connection between hidden nodes and output nodes. The task is accomplished in two phases. In first phase, connection between input nodes and hidden nodes are removed. For each ml w in the network, if

$$\max_p |v_{pm} w_{ml}| \leq 4\eta_2, \quad (2)$$

then remove ml w from the network. In the second phase, connections between hidden nodes and output nodes are removed. For each pm v in the network, if

$$|v_{pm}| \leq 4\eta_2 \leq 4\eta_2, \quad (3)$$

then remove pm v from the network.

- **Step 3** Remove connections between input nodes and hidden nodes further. If no weight satisfies condition (2) or condition (3), then for each ml w in the network,

$$w_{ml} = \max_p |v_{pm} w_{ml}| \quad (4)$$

Remove $ml w$ with smallest $ml w$. Continue, otherwise stop.

- **Step 4** Train again the network and calculating accuracy of the network in classification.
- **Step 5** If classification accuracy of the network falls below an appropriate level, then stop and use the previous setting of the network weights. Otherwise, head to **Step 2**.

6.3 (RE) Rule Extraction Algorithm[10,11]

Classification rules are sought in several areas from automatic knowledge acquisition to data mining and neural network rule extraction because some of their attractive options. They are understandable, explicit and verifiable by domain consultants, and may be modified, extended and passed on as standard knowledge. The rule extraction algorithm, will be applied to each numeric and discrete data, consist of three major functions:

a) Rule Extraction(RE): This function initialize the extracted rule list to be empty and sorts the examples according to example frequency. Then it picks the frequent occurring example as the base to generate a rule then it will add the rule to the list of extracted rules. Then it find all the example, that are covered by the rule and remove from the example space. It will repeats the above process iteratively and continuously adds the extracted rules to the rule list until the example space becomes empty.

b) Rule Clustering: The rules are clustered in terms of their category levels. Rules of the same category are clustered together as one group of rules.

c) Rule Pruning: Redundant(repeat) or more specific rules in each cluster are removed. In every clusters, more than one rule may cover the same example. For examples, the rule “if (color = green) and (height < 4) then grass” is already contained in a more general rule “if (color = green) then grass”, and thus the rule “if (color = green) and (height < 4) then grass” is redundant. Rule extraction eliminates these redundant rules in each cluster to further reduce the size of the best rule list.

The steps of the rule extraction(RE) algorithm are explained as follows:

- **Step 1** Extract Rule

The core of this step contains greedy algorithm that finds the shortest rule based on the primary order information, which may differentiate the pattern into consideration from the patterns of alternative classes. It then extracts shortest rules and take away the patterns covered by every rule until all patterns are coated by the rules.

- **Step 2** Cluster Rule: Cluster rules according to their category levels. Rules extracted in Step one are grouped in terms of their class levels.

- **Step 3 Prune Rule:** Replace specific rules with more general ones; Remove noise rules; Eliminate redundant rules;
- **Step 4** Check whether all patterns are coated by any principle on extraction. If affirmative then stop, otherwise continue.
- **Step 5** Determine a default rule on extraction. A default rule is chosen if no rule can be applied to a pattern.

7. PERFORMANCE EVALUATION[10,11]

This section evaluates the performance of the ESRNN algorithm on a set of well-known classification problems including diabetes, wine, iris that are widely used in data mining research and machine learning. The datasets representing all the issues were real world data.

7.1 Dataset Description

This section briefly describes the datasets utilized in this study. The datasets are summarized

Table 1. Characteristics of datasets.

SI No.	Datasets	No. of Examples	Input Attributes	Output Classes
1	Diabetes	768	8	2
2	Iris	150	4	3
3	Wine	178	13	3
4	Season	11	3	4
5	Golf Playing	14	4	2
6	Lenses	24	4	3

The diabetes dataset: The Pima Indians Diabetes information consists of 768 data pairs with eight attributes normalized between zero and one. The eight attributes are number of pregnancies (A1), plasma glucose concentration (A2), blood pressure (A3), triceps skin fold thickness (A4), Two hour serum insulin (A5), body mass index (A6), diabetes pedigree function (A7), and age (A8). In this database, 268 instances are positive (output equals 1) and 500 instances are negative (output equals 0).

The iris dataset: This is perhaps the best known database to be found within the pattern recognition literature. The set contains three classes of fifty instances each, where every class refers to a type of Iris plant. 4 attributes are used to predict the iris class, *i.e.*, sepal length (A1), sepal width (A2), petal length (A3), and petal width (A4), all in centimetres. Among the 3classes, class one is linearly separable from the other two classes, and classes two and three are not linearly separable from one another. To ease data extraction, we reformulate the data with three outputs, where class 1 is represented by {1, 0, 0}, class 2 by {0, 1, 0}, and class 3 by {0, 0, 1}.

The season data: The season dataset contains separate data only. There are eleven examples within the dataset, every of that consisted of three-elements. These are tree, weather and temperature. This was a four-class problem.

The golf playing data: The golf playing dataset contains both numeric and discrete data. There are 14 examples in the dataset, each of which consisted of four-elements. These are outlook, temperature, humidity and wind. This is a two-class problem.

7.2 Extracted Rules

The number of rules extracted by ESRNN algo. and the accuracy of the rules is Presented here in table.

Number of extracted rules and rules accuracies.

SI No.	Datasets	No. of Extracted Rules	Accuracy
1	Diabetes	2	76.56%
2	Iris	3	98.67%
3	Wine	3	91.01%
4	Season	4	100%
5	Golf Playing	3	100%
6	Lenses	8	100%

The diabetes data

Rule 1: If Plasma glucose concentration (A2) ≤ 0.64 and Age (A8) ≤ 0.69 then tested negative.
Default Rule: tested positive.

The iris data

Rule 1: If Petal-length (A3) ≤ 1.9 then iris setosa
Rule 2: If Petal-length (A3) ≤ 4.9 and Petal-width (A4) ≤ 1.6 then iris versicolor
Default Rule: iris virginica.

The season data

Rule 1: If Tree (A2) = yellow then autumn
Rule 2: If Tree (A2) = leafless then autumn
Rule 3: If Temperature(A3) = low then winter
Rule 4: If Temperature(A3) = high then summer
Default Rule: spring.

The golf playing data:

Rule 1: If Outlook (A1) = sunny and Humidity ≥ 85 then don't play
Rule 2: Outlook (A1) = rainy and Wind= strong then don't play
Default Rule: play.

7.3 Performance Comparisons[10]

This section compares experimental results of the ESRNN algorithm with the results of other works. The primary aim of this work is not to evaluate ESRNN in order to gain a deeper understanding of rule generation without an exhaustive comparison between ESRNN and all other works. Table 1 compares ESRNN results of the diabetes data with those produced by PMML, NN RULES, C4.5, NN-C4.5, OC1, and CART algorithms. ESRNN achieved 76.56% accuracy although NN-C4.5 was closest second with 76.4% accuracy. Due to the high noise level, the diabetes problem is one of the most challenging problems in our experiments. ESRNN has outperformed all other algorithms.

Performance comparison of the ESRNN with other algorithms for the diabetes data.

Dataset	Feature	ESRNN	PMML	NN RULES	C4.5	NN-C4.5	OC1	CART
Diabetes	No. of Rules	2	2	4	–	–	–	–
	Avg. No. of Conditions	2	1	3	–	–	–	–
	Accuracy (%)	76.56	75	76.32	70.9	76.4	72.4	72.4

Table 2 compares ESRNN results of the iris data with those produced by PMML , NN RULES , DT RULES , BIO RE , Partial RE , and Full RE algorithms. ESRNN achieved 98.67% accuracy although NN RULES was closest second with 97.33%accuracy. Here number of rules extracted by ESRNN and NN RULES are equal.

Performance comparison of the ESRNN algorithm with other algorithms for the irish data.

Dataset	Feature	ESRNN	PMML	NN RULES	DT RULES	BIO RE	Partial RE	Full RE
Irish	No. of Rules	3	3	3	4	4	6	3
	Avg. No. of Conditions	1	1	1	1	3	3	2
	Accuracy (%)	98.67	91.3	97.33	94.67	78.67	78.67	97.33

Table 3 compares the ESRNN results of the season data with those produced by RULES and X2R . All three algorithms achieved 100% accuracy. This is possible because the number of examples is low. ESRNN extracted five rules, whereas RULES extracted seven and X2R six.

Performance comparison of ESRNN with other algorithms for season data.

Dataset	Feature	ESRNN	RULES	X2R
Season	No. of Rules	5	7	6
	Avg. No. of Conditions	1	2	1
	Accuracy (%)	100	100	100

Table 4 compares ESRNN results of golf playing data with those produced by RULES , RULES-2 , and X2R [25]. All four algorithms achieved 100% accuracy because the lower number of examples. Number of extracted rules by ESRNN are 3 whereas these were 8 for RULES and14 for RULES-2.

Performance comparison of ESRNN with other algorithms for golf playing data.

Dataset	Feature	ESRNN	RULES	RULES-2	X2R
Golf Playing	No. of Rules	3	8	14	3
	Avg. No. of Conditions	2	2	2	2
	Accuracy (%)	100	100	100	100

8. CONCLUSION

In this paper, We present research on data mining based on neural network. At present, data mining is a new and important area of research, and neural network itself is very suitable for solving the problems of data mining because its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage, high degree of fault tolerance & network structure i.e. recurrent network. The combination of data mining and neural network can greatly improve the efficiency of data mining, and it has been widely used & we have presented

neural network based data mining scheme to mining classification rules from given databases. This work is an attempt to apply the approach to data mining by extracting symbolic rules. An important feature of the rule extraction algorithm is its recursive nature. A set of experiments was conducted to test the approach using a well defined set of data mining problems. The results indicate that, using the approach, high quality rules can be discovered from the given data sets. The extracted rules are concise, comprehensible, order insensitive, and do not involve any weight values. The accuracy of the rules from the pruned network is as high as the accuracy of the fully connected networks. Experiments showed that this method helped a lot to reduce the number of rules significantly without sacrificing classification accuracy. In almost all cases ESRNN outperformed the others. With the rules extracted by the method here, ANNs should no longer be regarded as black boxes. Since, black boxes are diminished & more researchers use them. Thus, neural networks are becoming very popular with data mining practitioners.

REFERENCES

- [1] M. Charles Arockiaraj “Applications of Neural Networks In Data Mining”, Arakkonam, (Research Inveny: International Journal Of Engineering And Science Vol.3, Issue1),May 2013.
- [2] Dr. Yashpal Singh ,Alok Singh Chauhan “Neural Networks In Data Mining”, India, (Journal of Theoretical and Applied Information Technology)2005.
- [3] K. Amarendra, K.V. Lakshmi & K.V. Ramani “Research on Data Mining Using Neural Networks” , India
- [4] Xianjun Ni “Research of Data Mining based on Neural Networks” ,China , (World Academy of Science, Engineering and Technology Vol:2) ,2008.
- [5] Sonalkadu, Prof.Sheetal Dhande “Effective Data Mining Through Neural Network”, (International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 3) ,March 2012
- [6] Vidushi Sharma ,Sachin Rai ,Anurag Dev “A Comprehensive Study of Artificial Neural Networks”, India (International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10) ,October 2012
- [7] Ms. Sonali. B. Maind ,Ms. Priyanka Wankar “Research Paper on Basic of Artificial Neural Network”, Wardha ,(International Journal on Recent and Innovation Trends in Computing and Communication Volume: 2 Issue: 1),January 2014.
- [8] Anil K. Jain, Jianchang Mao, K.M. Mohiuddin “Artificial Neural Networks : A Tutorial” , Michigan ,March 1996
- [9] Ajith Abraham “Artificial Neural Networks” Oklahoma State University, Stillwater, USA 2005.
- [10] S. M. Kamruzzaman and A. M. Jehad Sarkar “A New Data Mining Scheme Using Artificial Neural Networks”, Korea , 28 April 2011.
- [11] Mrs.Maruthaveni.R, Mrs. Renuka Devi.S.V ” Efficient Data Mining For Mining Classification Using Neural Network”(International Journal Of Engineering And Computer Science Volume 3 Issue 2) February , 2014.

AUTHORS

The author Gaurab Tewary is an MCA from Northern India Engineering College, New Delhi. Under GGSIP University, New Delhi.

