**DATA MINING USING NEURAL NETWORKS**

**1)Dr. G.R.Bamnote**

**2)Mr.S.R.Patil**

**1) Head Of Dept.PRMIT&R, Badnera. 2) M.E. II Sem FT.PRMITE, Bandera.**

**ABSTRACT :**

**Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.**

**.**

**INTRODUCTION:**

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data.

Consider the 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.

Another example of where this institution has failed

to utilize its 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.

These are simple examples of what could be

achieved using data mining.

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.

**2. ARTIFICIAL NEURAL NETWORKS:**

An **artificial neural network** (ANN), often just

called a "neural network" (NN), is a mathematical

model or computational model based on biological

neural networks, in other words, is an emulation of

biological neural system. It consists of an

interconnected group of artificial neurons and

processes information using a connectionist

approach to computation. In most cases an ANN is

an adaptive system that changes its structure based

on external or internal information that flows

through the network during the learning phase

****

**2.1 Neural Network Topologies:**

**Feedforward neural network:** The feedforward

neural network was the first and arguably simplest

type of artificial neural network devised. In this

network, the information moves in only one

direction, forward, from the input nodes, through

the hidden nodes (if any) and to the output nodes.

There are no cycles or loops in the network. The

data processing can extend over multiple (layers of)

units, but no feedback connections are present, that

is, connections extending from outputs of units to

inputs of units in the same layer or previous layers.

**Recurrent network:** Recurrent neural networks

that do contain feedback connections. Contrary to

feedforward networks, recurrent neural networks

(RNs) are models with bi-directional data flow.

While a feedforward network propagates data

linearly from input to output, RNs also propagate

data from later processing stages to earlier stages.

**2.2 Training Of Artificial Neural Networks:**

A **neural network** has to be configured such that

the application of a set of inputs produces (either

'direct' or via a relaxation process) the desired set of

outputs. Various methods to set the strengths of the

connections exist. One way is to set the weights

explicitly, using a priori knowledge. Another way is

to **'train' the neural network** by feeding it

teaching patterns and letting it change its weights

according to some learning rule. We can categorize

the learning situations as follows:

• **Supervised learning** or Associative learning

in which the network is trained by providing it

with input and matching output patterns. These

input-output pairs can be provided by an external teacher, or by the system which

contains the neural network (self-supervised).

• **Unsupervised learning** or Self-organization in

which an (output) unit is trained to respond to

clusters of pattern within the input. In this

paradigm the system is supposed to discover

statistically salient features of the input

population. Unlike the supervised learning

paradigm, there is no a priori set of categories

into which the patterns are to be classified;

rather the system must develop its own

representation of the input stimuli.

**Reinforcement Learning** This type of

learning may be considered as an intermediate

form of the above two types of learning. Here

the learning machine does some action on the

environment and gets a feedback response

from the environment. The learning system

grades its action good (rewarding) or bad

(punishable) based on the environmental

response and accordingly adjusts its

parameters.

**3. 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.

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

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

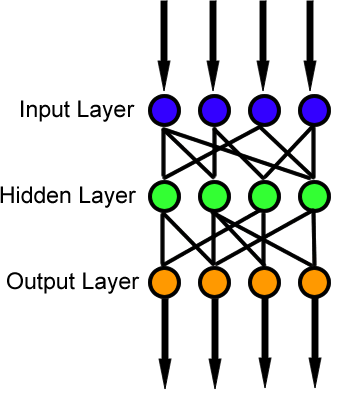
neural network.



**3.1. Feedforward Neural Network**:

A feedforward neural network is an artificial neural network where connections between the units do not form a directed cycle. This is different from recurrent neural networks.

The feedforward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.



The simplified process for training a FFNN is as

follows:

1. Input data is presented to the network and

propagated through the network until it reaches

the output layer. This forward process produces

a predicted output.

2. The predicted output is subtracted from the

actual output and an error value for the

networks is calculated.

3. The neural network then uses supervised

learning, which in most cases is back

propagation, to train the network. Back

propagation is a learning algorithm for

adjusting the weights. It starts with the weights

between the output layer PE’s and the last

hidden layer PE’s and works backwards

through the network.

4. Once back propagation has finished, the

forward process starts again, and this cycle is

continued until the error between predicted and

actual outputs is minimized.

**3.2. The Back Propagation Algorithm:**

**Backpropagation**, or **propagation of error**, is a

common method of teaching artificial neural

networks how to perform a given task.The back

propagation algorithm is used in layered feedforward

ANNs. This means that the artificial

neurons are organized in layers, and send their

signals “forward”, and then the errors are

propagated backwards. The back propagation

algorithm uses supervised learning, which means

that we provide the algorithm with examples of the

inputs and outputs we want the network to

compute, and then the error (difference between

actual and expected results) is calculated. The idea

of the back propagation algorithm is to reduce this

error, until the ANN *learns* the training data.

**Summary of the technique:**

1. Present a training sample to the neural

network.

2. Compare the network's output to the desired

output from that sample. Calculate the error in

each output neuron.

3. For each neuron, calculate what the output

should have been, and a *scaling factor*, how

much lower or higher the output must be

adjusted to match the desired output. This is

the local error.

4. Adjust the weights of each neuron to lower the

local error.

5. Assign "blame" for the local error to neurons at

the previous level, giving greater responsibility

to neurons connected by stronger weights.

6. Repeat the steps above on the neurons at the

previous level, using each one's "blame" as its

error.

**Accounting**

Identifying tax fraud

Enhancing auditing by finding irregularities

**Finance**

Signature and bank note verificatio

Risk Management

Foreign exchange rate forecasting

Bankruptcy prediction

Customer credit scoringCredit 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.

**6. DESIGN PROBLEMS:**

There are no general methods to determine the

optimal number of neurones necessary for

solving any problem.

It is difficult to select a training data set which

fully describes the problem to be solved.

**SOLUTIONS TO IMPROVE ANN**

**PERFORMANCE:**

Designing Neural Networks using Genetic

Algorithms

Neuro-Fuzzy Systems

**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 indata

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 45, no. 8

[7] Fausett, Laurene (1994), Fundamentals of

Neural Networks: Architectures, Algorithms

and Applications, Prentice-Hall, New Jersey,

USA.

[8] <http://ieeexplore.ieee.org/stamp>

[9] <http://www4.rgu.ac.uk/files/chapter3%20-%20bp.pdf>

[10]https:dspace.ist.utl.pt/bitstream/2295/57975/1/licao\_20.pdf

[11] <http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=250074&abstractAccess=no&userType=inst>

[12]<http://airccse.org/journal/ijdkp/papers/2512ijdkp02.pdf>