

# Accord of ANN in Data Mining

Arindam Giri

Haldia Institute of Technology.  
Department of Computer Science  
and Engineering  
Haldia, WB, India

Debaprasad Misra

Haldia Institute of Technology.  
Department of Computer Science  
and Engineering  
Haldia, WB, India

Manas Kumar Ray

National Institute Of Technology,  
Durgapur  
Department of Computer Science  
And Engineering W.B,India

## ABSTRACT

The joint venture of artificial neural network and data mining make the process more perfect, powerful, fast, distributed, fault and noise tolerance and independence of prior assumption. This paper is an overview of artificial neural network, data mining, knowledge management and the purpose of ANN in the field of data mining. Due to the huge amount of data in the data warehouse and data bases, companies try to find the actual and perfect data (values) for making the "knowledge" which helps to implement the appropriate and effective decision making for their company. Data mining is an area that, extract the hidden predictive information from large data storage area with its powerful technology. The main objectives of data mining are –classification, clustering, association rule evaluation learning, and regression. DM is the business of answering question that 'you have not asked yet'. The artificial neural networks (ANN), among different soft computing methodologies are widely used to meet the challenges thrown by data mining due to their robust, powerful, distributed, fault tolerant computing and capability to learn in a data-rich environment.

## General Terms

Artificial Neural Network, Data Mining

## Keywords

ANN, Data mining, KDD and KM, Features and application of ANN, Accord of ANN in DM

## 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. Data mining can be viewed as a result of the natural appraisal of information or values. The information and knowledge can be used for claim ranging from market analysis, fraud detection and customer actions. Due to the significance of data ware house, data mining introduce the advance data analysis including the 'knowledge' making which helps to implement the decision making for the companies. It is a business of answering that 'what you have not asked yet'.

Neural network is a non linear, parallel, distributed, fault and noise tolerance and independence of prior assumption network. It is based on the human nervous system, according to the features of biological neurons by simplifying, summarizing and verifying.

The high affordability to the to the noise data, low error rate, network training algorithm, rule evaluation and extracting methods makes neural network more popular and powerful tool for the data mining users.

## 2. DATA MINING

In the mid of 1980's data warehouse used for decision support system for many companies. Decision support system (DSS), Artificial Intelligence (AI), and Computer Science (CS) makes the combinational approaches for many companies to support and cultivate their actual and perfect data.

Data mining is an area of computer science that focuses the hidden analytical information to make the 'knowledge' and helps to build up the DSS. The main six key factors makes DM more powerful and perfect, these are –

1. High quality data,
2. The right tool,
3. An adequate sample size,
4. The right data, extract and
5. Transform data rate and
6. Load data in the data bases.

The main goals of DM included

1. Regression.
2. Classification.
3. Clustering.
4. Association rule evaluation for learning
5. Prediction.

Data mining tasks are categorized into two parts –descriptive and predictive.

Descriptive DM focused on provides information to understand what is happening inside the data without predetermined idea. For example- PR, association rules.

Predictive DM allows the user to submit records with unknown field values, and the system will gives the unknown values based on previous pattern discovered from database. Examples are – classification, prediction.

The methods are mainly used for gaining knowledge in data mining are as follows

1. Statistical methods (prediction of time series, cluster analysis etc.)
2. Production rules IF . . . THEN
3. Decision trees
4. Genetic algorithms
5. Neural networks

The architecture of DM system is given below,

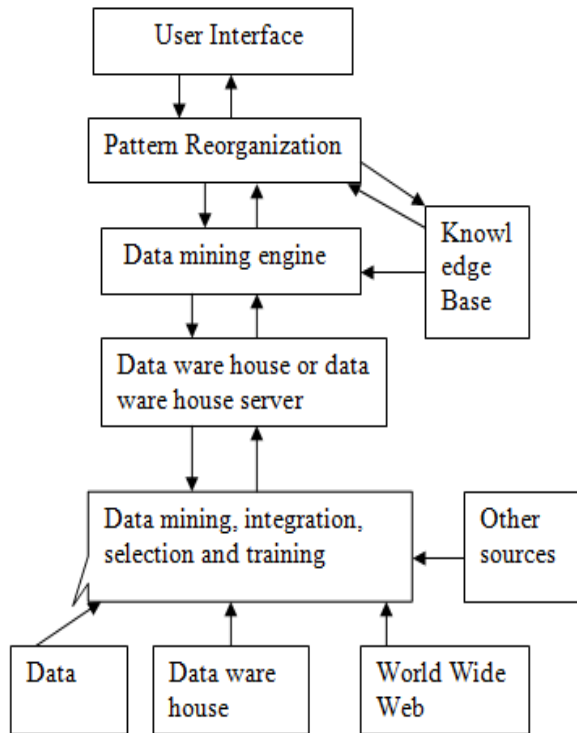


Fig -1, DM architecture

### 3. APPLICATION OF DATA MINING

Today's the cosmic album and store of data are very crucial and significant for the companies. As consequence of applied techniques of DM in the data storage area are helpful for making 'knowledge' which leads to decision making for particular application. DM use in the all parts of data bases, data warehouses and decision making part for a company. In a summarized form the application of DM as follows:-

Industry	Application
Finance	Credit Card Analysis
Insurance	Claims, Fraud Analysis
Telecommunication	Call record analysis
Transport	Logistics management
Data Service providers	Value added data
Utilities	Power usage analysis
Consumer goods	Promotion analysis

Table1 – Application of DM

### 4. ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial Neural Network (ANN) is physical cellular system which can acquire, store and utilized experimental knowledge (according to Jurada J.M (1992)). The term was coined by McCulloch and Pitts (1943). According to Haykin, the NN resemble the brain in two respect –

1) Knowledge is acquired by the neural network through a learning process.

2) Interneuron connection strength known as synaptic weight are used to store the knowledge.

The invention and development of ANN has some focused, these included,

- 1) The neurons or nodes or units or cells are simple elements for information transforming.
- 2) Signals help to maintain the link between the neurons.
- 3) Associated weight measured in the neural net which multiple the signal transmitted.
- 4) Non linear activation function used in the each neuron to its net input to calculate the desired output signal.

The main three parts of ANN are –

- 1) The architecture or model
- 2) The learning algorithm or training and
- 3) The activation function.

The activation function is any real input range between 0 to 1 and -1 to 1. Some of the popular activation functions are Linear or identity  $f(x) = x$

Hyperbola tangent  $f(x) = \tanh(x)$

Threshold  $f(x) = 0$  if  $x < 0$ , 1 otherwise.

The main eight steps for neural network application development are-

1. Data collection,
2. Training and testing data,
3. Separation,
4. Network architecture,
5. Parameter tuning and weight utilization,
6. Data transformation training,
7. Testing and
8. Implementation.

The various methods for learning in Neural Networks are basically three types – supervised, unsupervised and reinforced learning. Apart from these other methods are Hebbin learning, complimentary learning, stochastic learning.

### 5. NEURAL NETWORK FEATURES

The main features of ANN are as given below:-

1. Neural Network is composed of a large number of very simple processing elements called neurons.
2. Each neuron is connected to other neurons by means of connections or links with an associated weight.
3. Memories are stored or represented in a neural network in the pattern of interconnection strengths among the neurons.
4. Information is processed by changing the strengths of interconnections and/or changing the state of each neuron.
5. A neural network is trained rather than programmed.
6. A neural network acts as an associative memory. It stores information by associating it with other information in the memory.

### 6. NEURAL NETWORK APPLICATIONS

The tasks artificial neural networks are applied to tend to plunge within the following broad categories:

1. Function approximation, or regression analysis, including time series prediction, fitness approximation and modeling.
2. Classification, including pattern and sequence recognition, novelty detection and sequential decision making.
3. Data processing, including filtering, clustering, blind source separation and compression.
4. Robotics, including directing manipulators, Computer numerical control.

Make use of areas include system identification and control (vehicle control, process control), quantum chemistry, game-playing and decision making (backgammon, chess, poker), pattern recognition (radar systems, face identification, object recognition and more), sequence recognition (gesture, speech, handwritten text recognition), medical diagnosis, financial applications (automated

trading systems), data mining (or knowledge discovery in databases, "KDD"), visualization and e-mail spam filtering.

## 7. KDD, KM AND DATA MINING

There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data. *Knowledge Discovery in Databases (KDD)* is the process of automatic discovery of previously unknown patterns, rules, and other regular contents implicitly present in large volumes of data. Data Mining (DM) denotes discovery of patterns in a data set previously prepared in a specific way. DM is often used as a synonym for KDD. However, harshly speaking DM is just a central phase of the entire process of KDD. Knowledge discovery is the process of nontrivial extraction of information from data, information that is implicitly present in that data, previously unknown and potentially useful for the user. The information must be in the form of patterns that are comprehensible to the user (such as, for example, If-Then rules).

For a data set  $F$ , represented in a language  $L$ , if  $FS$  is a subset of  $F$  and  $c$  denotes certainty measure, then a pattern is an expression  $S$  in the language  $L$  with the certainty  $c$  about some relation(s) among the data in  $F$ s. In order for the expression  $S$  to really be a pattern, it must be simpler than merely counting all the data in  $F$ .

Knowledge Management (KM) is to discover, develop, utilize, deliver, and absorb knowledge inside and outside the organization through an appropriate management process to meet current and future needs according to Quintas et al. (1997) and it is the creation, extraction, transformation and storage of the correct knowledge and information in order to design better policy, modify action and deliver results. The main success factors focused in KM are as follows:-

1. A trusting and open organizational
2. Culture
3. Senior management leadership and commitment
4. Employee involvement
5. Employee training
6. Trustworthy teamwork
7. Employee empowerment
8. Information systems infrastructure etc.

Data mining and knowledge discovery have been topics considered at many AI, database and statistical approaches. The purpose is that, gradually introduce the process of KDD and typical DM tasks. The idea of automatic knowledge

discovery in large databases is first presented informally, by describing some practical needs of users of modern database systems. Several important concepts are then formally defined and the typical context and resources for KDD. The scope of KDD and DM is in terms of classification of KDD/DM problems and common points between KDD and several other scientific and technical disciplines that have well-developed methodologies and techniques used in the field of KDD. It describes the typical KDD process, DM tasks and some algorithms that are most frequently used to carry out such tasks. Some other important aspects of KDD, such as using domain knowledge in the KDD process and evaluating discovered patterns. Finally we can say that, important of KDD application domains and practical KDD/DM systems makes several hot topics and research problems in the field that are of interest to companies for real time application.

## 8. ACCORD OF ANN IN DATA MINING

Due to some important characteristic of ANN such as fast, robustness, powerful, independence of prior assumption and classification making the wide range of role in the data mining field. The knowledge discovery in database (KDD) with the help of some data extracting method makes the joined operation performed by ANN and data mining successfully. The data mining based on neural network constructed by

- 1) Data preparation (It includes data cleaning, data options, data preprocessing and data expression)
- 2) Rule extracting (by using algorithm and training).
- 3) Rule assessment.
- 4) Gaining knowledge and
- 5) Make DSS (by using some specific tools).

The main two types of data mining and neural network based networks are

Data mining based on self organization neural network-

The Finland scholars T. Kohonen proposed a model related to human brain working process. For example Self Organization Feature Map (SOFM). Based on competitive learning (Unsupervised). Only one output neuron activated at any one time .Winner-takes-all neuron or winning neuron. In a Self-Organizing Map Neurons placed at the nodes of a lattice one or two dimensional. Neurons selectively tuned to input patterns by competitive learning process locations of neurons so tuned to be ordered formation of topographic map of input patterns. Spatial locations of the neurons in the lattice, intrinsic, statistical features contained in the input patterns

Data mining based on fuzzy neural network-

Make the output intuitively and more stable fuzzy logic applied. Some examples are Fuzzy perception model, Fuzzy BP model, Fuzzy interface network etc.

The synapse resistance to the incoming signal can be changed during learning process. Data mining aspirants now a day improves some algorithm based methodology for interpret the process in the real world application.

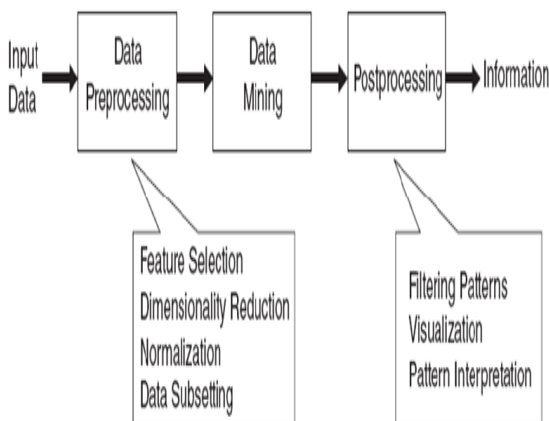


Fig 2 The process of KDD

## 9. CONCLUSIONS

So the data warehouse communal are now focused on the poison of ANN in the data mining area, which makes the path to lead the fast, active, robust and powerful process for entire application for a company. Data mining based on ANN is an important tactic and course of action that in communal the business world for their making 'knowledge' as well as help to build the decision making to their company. Now, days the data mining makes its own image for classification and pattern reorganization in the entire data ware houses and the neural networks perform the 'catalyst' part on the process. The key technique of implementation the two techniques as single approach we have to focused on the actual data, knowledge, the model and input output interface. So, ultimately the operation make successful, robust, fast, fault and noise tolerance as consequence of the joint venture of data mining and artificial neural nets.

## 10. REFERENCE

- [1] Dr Yashpal Singh, Alok Singh Chauhan . Neural Networks In Data Mining, Journal of Theoretical and Applied information Technology, 2005, p 37 to 42
- [2] Xianjun Ni , Research of Data Mining Based on Neural Networks, World Academy of Science, Engineering and Technology ,39,2008, p 381-384
- [3] Cynnthia Krieger , Neural Networks in Data Mining ,1996
- [4] Dr Hamid Nemati, Introduction to Data Mining Using ANN,ISM 611
- [5] Ms. Smita Nirkhi, Potential use of ANN in Data Minng, 978-1-4244-5586-7/10, IEEE 2010
- [6] Han and Kamber, Data Mining Concept and Technique, 2010,Morgan Kaufmann Publishers
- [7] S.Haykin, Neural Networks and Learning Machines,2010,PHI
- [8] Data Mining: Concepts and Techniques Jiawei Han and Micheline Kamber, Morgan Kaufmann, 2001
- [9] A.Vesely "Neural Networks In Data Mining" Agric. Econ. – Czech, 49, 2003 (9): 427–431
- [10] Introduction to Artificial Neural Networks. Nicolas Galoppo von Borries. [www.cs.unc.edu/~nico/courses/comp290-58/nn-presentation/ann-intro.pdf](http://www.cs.unc.edu/~nico/courses/comp290-58/nn-presentation/ann-intro.pdf)
- [11] Knowledge Discovery and Data Mining in Databases Vladan Devedzic FON - School of Business Administration, University of Belgrade, Yugoslavia

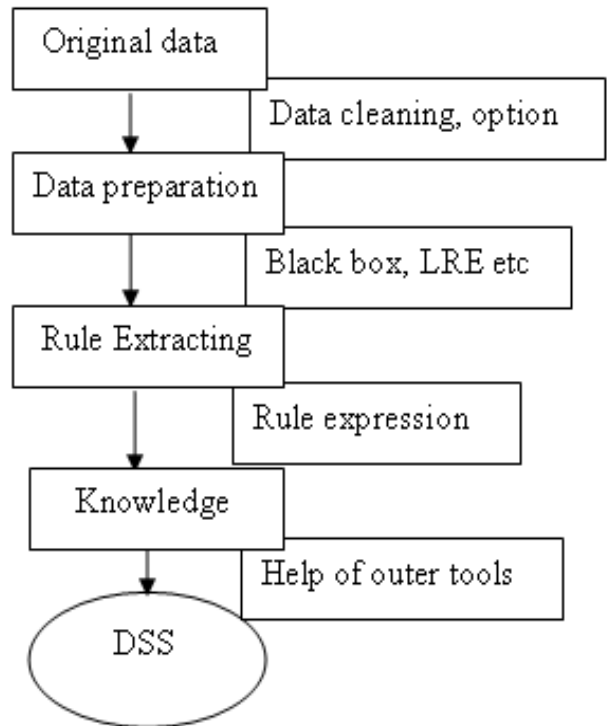


Fig 3 Data mining based ANN