International Journal of Computer Science and Software Engineering
Volume 3, Number 1 (2017), pp. 1-9

http://www.irphouse.com

# Neuro-fuzzy, GA-Fuzzy, Neural-Fuzzy-GA: A Data Mining Technique for Optimization

**Dr. Anjali B. Raut**
*H.O.D.*
*Computer Science and Engineering Department*
*H.V.P.M's College of Engineering and Technology*
*Amravati, India.*

**Abstract**

Now a day's data mining is very thirsty area for researchers. For the extraction of the hidden predictive information from large databases, data mining is a powerful new technology having great potential to analyze important information from large data. Various algorithms have been proposed in the past for the mining process out of which neural based mining algorithms are predominant. This paper focuses on work utilization of a genetic optimization algorithm for minimizing and bringing about an optimal search value. This paper presents an elaborative review of types and existing techniques.

**Keywords:** Algorithm, Data mining, neural network, genetic optimization, KDD knowledge discovery database, data warehouse.

## I. INTRODUCTION

The computerization of our society has substantially increased our capabilities for both generating and collecting data from different sources. A tremendous amount of data has torrent almost every aspect of our lives. The amount of data stored on web is growing very fast and for this data manual analysis, cannot be possible. Thus for the automated extraction of patterns representing knowledge implicitly stored the data mining (DM), also known as knowledge discovery from data (KDD), which is used.

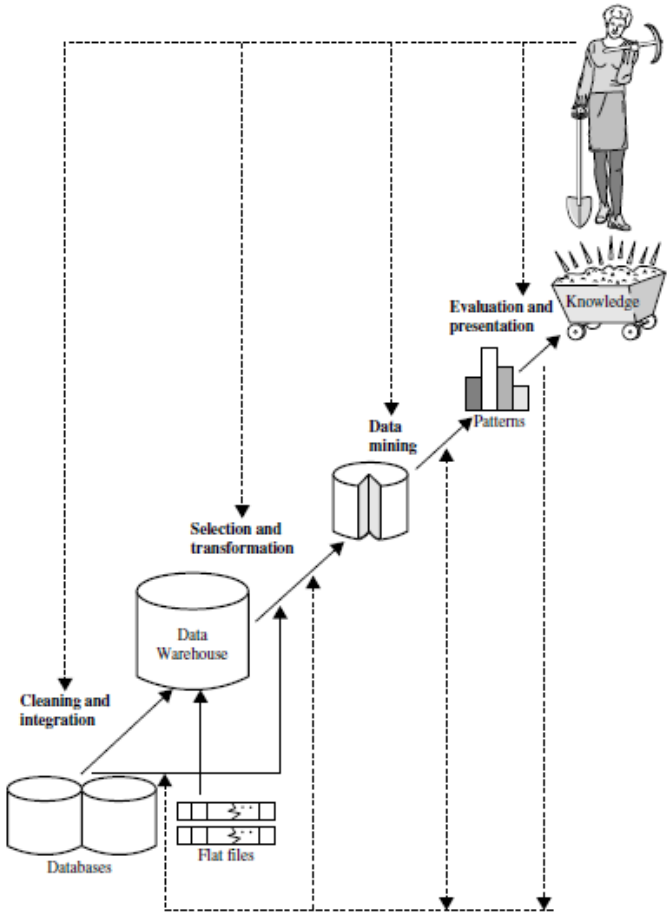


**Figure 1:** Data mining as a step in process knowledge discovery.

Figure 1 shows the knowledge discovery process as an iterative sequence of following steps.

1. Data Cleaning
2. Data integration
3. Data selection
4. Data Transformation
5. Data mining
6. Pattern evaluation
7. Knowledge presentation

Step 1 to 4 may interact with the user or a knowledge base. In a general data mining system which consist of the knowledge base and the data base. Data cleaning is the first and foremost step of an iterative data mining process. It is similar to removing of the noisy, redundant and irrelevant data from the data set. This step is followed by data integration where the data from multiple datasets are combined into a common

source. In data selection data relevant to the analysis task are retrieved from the databases which are followed by the data transformation step. To extract data patterns, an essential process or intelligent methods are applied under the data mining stage. Which is further followed by pattern evaluation and knowledge presentation stages.

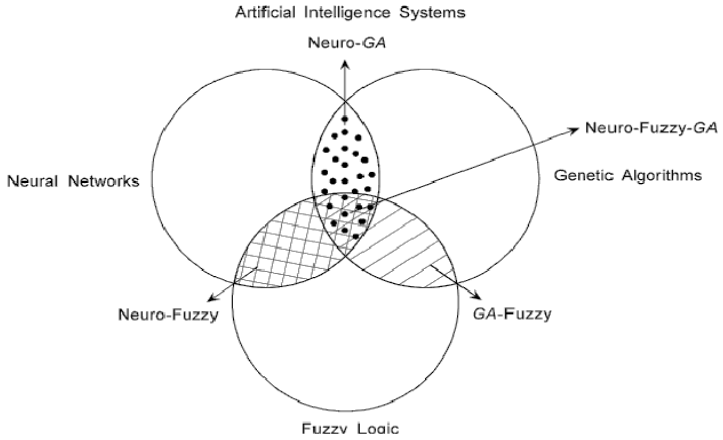


**Figure 2:** Integration of neural networks, fuzzy logic, and genetic algorithm technologies.

In this paper, three technologies were focused, namely Neural Network (NN), Fuzzy Logic (FL), and Genetic Algorithms (GA) and their hybrid combinations. As illustrated in above Figure 2, each of these technologies individually and in combination are used to solve problems. The different combinations are *neuro-fuzzy, GA-fuzzy, neural-fuzzy-GA* technologies.

## II. LITERATURE REVIEW

Data mining which is referred as knowledge discovery in databases is a process of extraction of implicit, previously unknown and potentiality useful information from large data sets [1]. The extracted information is also referred as knowledge of the form rules, constraints and regularities. For rule mining, researchers have been using many techniques such as decision tree, database, statistical, AI, cognitive etc. Various data mining methods, including characterization, generalization, classification, clustering, association, evolution, pattern matching, have been reviewed in [12].

One of the best known data mining techniques is association. In association, based on a relationship between items in the same transaction sequential patterns are discovered [17]. Therefore the association technique is also known as relation technique. The mining of association rules is one of the most popular solution to find interesting patterns from databases such as association rules, correlations, sequences,

classifiers, clusters etc. [3] [4]. Sequential patterns analysis [8] is one of data mining technique that works to discover or identify related patterns, regular events or trends in transaction data over a business period. Rule mining using neural networks (NNs) is the one of the challenging task as there is no straight way to translate NN weights to rules.

A data warehouse is supported to be a place where data gets stored so that applications can access and share it easily. The type of industry and the company will decide the type of stored data. Classification technique [7] [13] classify each item in a set of data into one of predefined set of classes. Classification technique uses decision trees, neural network, and statistics. A Bayesian classifier is used internally, the conventional support vector machine classifier, back propagation classifier whose methodology and implementation details has been discussed by the author [14].

For classification artificial neural networks (ANNs) are used. They are efficient for finding commonalities in a set of unrelated data so that preferred in number of classification tasks. But problem with ANNs when used with classification is that, while a trained ANN [11] can indeed classify the data, sometimes with more accuracy than a traditional, symbolic machine learning approach, the reasons for their classification cannot be found without difficulty. Trained ANNs are commonly perceived to be dark box which map input data onto a class through a number of mathematically weighted connections between neurons.

The genetic algorithm has some benefits:

- It can solve every optimization problem which can be described with the chromosome encoding.
- the genetic algorithm cannot make many mathematical requests to the optimization;
- Compared with the local search of tradition, the evolution operation of genetic algorithm enables it to carry on the effective global optimization.
- The genetic algorithm provided very big flexibility to deal with various domains overlapping issues.

The process of the genetic algorithm is mainly: 1st, basis domain of definition and needs the precision to carry on the code; 2nd, initialization population; 3rd, in the sufficiency of individual to the population appraises; 4th, the sufficiency of basis individual, the choice individual carries on the hybrid, variation and some other genetic manipulations, produces the next generation population; 5th, duplicates 2-4 processes, until finding the most appropriate solution.

*1) Neural Network*

A neural network is a highly interconnected network of a large number of processing elements called neurons. A neural network is characterized by

- Its pattern of connections between the neurons called its architecture
- Its method of determining the weight on the connections i.e. supervised or unsupervised and
- Its activation function.

The back propagation algorithm in which synaptic strengths are systematically modified so that the response of the network increasingly approximates the desired response is considered an optimization problem. The computed output is compared with the actual input. The difference is used to update the weights of each layer i.e. input layer, hidden layer, output layer using the delta rule [6, 9].

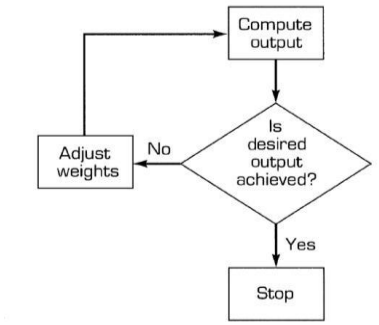


**Figure 3:** flow of weight update

*2) Artificial Neural Network*

The human brain no doubt is a highly complex structure viewed as a massive, highly interconnected network of simple processing elements called neurons. Artificial neuron is a structural model bases on the human brain. Basically an engineering approach of biological neuron is an Artificial Neuron. ANN is defined as a data processing system consisting of large number of simple highly interconnected processing elements in an architecture inspired by the structure of brain.

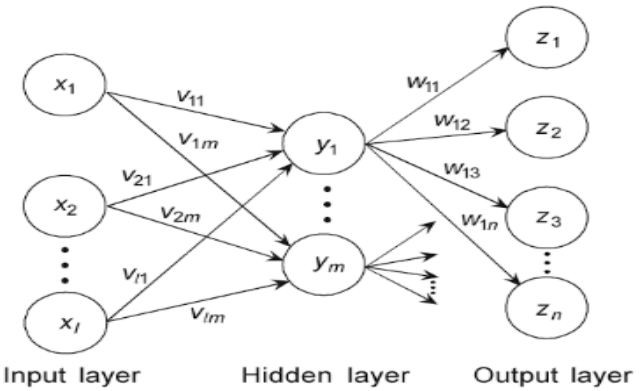


**Figure 4:** A multi layered feed forward architecture

The feed-forward neural network architecture as indicates is made up of multiple layers. Thus, architectures of this class besides processing an input and an output layer also have one or more intermediary layers called hidden layers. The computational units of the hidden layer are known as the hidden neurons. The hidden layer aids in performing useful intermediary computations before directing the input to the output layer. The input layer neurons are linked to the hidden layer neurons and the weights o these links are referred to as input-hidden layer weights. Feed-forward networks are often trained using a back propagation-learning scheme. Back propagation learning works by making modifications in weight values starting at the output layer then moving backward through the hidden layers of the network.

*1) Input: Data set*
*2) Target: Classified set*
*3) Generate m x n map with a seed neurons*
*4) Initialize initial weight W(0)*
*Select an instance n*
*Find the winning neuron*
*Determine the error*
*5) Adapt the weight vectors of the k neuron using genetic optimization after encoding*
*Repeat steps until convergence*
*Add or delete connections between neurons according to the measure distances*

*3) Fuzzy Logic*

Logic is the science of reasoning. Symbolic or mathematical logic has turned out to be a powerful computational paradigm. Fuzzy logic starts with a set of user-supplied human language rules. The fuzzy systems convert these rules to their mathematical equivalents. This simplifies the job of the system designer and the computer, and results in much more accurate representations of the way systems behave in the real world. A paradigm is a set of rules and regulations which defines boundaries and tells us what to do to be successful in solving problems within these boundaries. For example the use of transistors instead of vacuum tubes is a paradigm shift - likewise the development of Fuzzy Set Theory from conventional bivalent set theory is a paradigm shift.

*4) GA Optimization*

Genetic algorithms are good at taking larger, potentially huge, search spaces and navigating them looking for optimal combination of things and solutions which we might not find in a life time.

The genetic algorithms are search algorithms based on the mechanism of natural selection and natural genetics. The GA is basically based on biological principle of

natural selection. The architecture of systems that implement GAs is able to adapt to a wide range of problems. When using GA to solve a problem, the first thing, and perhaps the most difficult task that must be determined is how to model the problem as a set of individuals. In GA, reproduction is defined by precise algorithms that indicate how to combine the given set of individuals to produce new ones. These are called crossover. It then evaluates each of those solutions, and decides on a fitness level for each solution set. These solutions then breed new solutions. The parent solutions that are more fit are more likely to reproduce, while those that are less fit are more unlikely to do so. In essence, solutions are evolved over time. This way the search space evolves to reach the point of the solution. The GAs having three elements that are the encoding, the operator and the fitness function. The individuals in genetic space are chromosomes. The basic constitution factors are genes. The position of gene in individual is called locus. A set of individuals constructs a population. The fitness represents the evaluation of adaptability of individual to environment.

The elementary operation of genetic algorithm consists of three operands: selection, crossover and mutation. Select is also called copy or reproduction. By calculating the fitness $f_i$ of individuals, it selects high quality individuals with high fitness, copy them to the new population and eliminate the individual with low fitness to generate the new population. Generally used strategies of selection include roulette wheel selection, expectation value selection, paired competition selection and retaining high quality individual selection. Crossover puts individuals in population after selection into match pool and randomly makes individuals in pairs to form parent generation. Then according to crossover probability and the specified method of crossover, it exchanges part of the genes of individuals that is in pairs to form new pairs of child generation and finally to generate new individuals. Generally used methods of crossover are one point crossover, multi point crossover and average crossover. According to specified mutation rate, mutation substitutes genes with their opposite genes in some loci to generate new individuals.

## III. CONCLUSION

A research on data mining based on neural network optimized through genetic algorithm is presented in this paper. Data mining is known for its high robustness, self organizing adaptive, parallel processing capabilities, and distributed storage with a high degree of fault tolerance. The combination of data mining , neural network and genetic algorithm can greatly improve the efficiency of data mining, and it has been widely used & has presented neural network based data mining scheme to mining classification rules from given databases. An important feature of the rule extraction algorithm is its recursive nature.

**REFERENCES**

[1] M. Charles Arockiaraj "Applications of Neural Networks In Data Mining", International Journal Of Engineering And Science Vol.3, Issue. 1, 2013.

[2] Sachin Sharma and Savita Shiwani, "Data mining based accuracy enhancement of ANN using Swarm intelligence", International journal of communication and computer technologies, Vol. 2, No. 9, 2014.

[3] Bhatia and Jyoti, "An Analysis of heart disease prediction using different data mining techniques, International Journal of Engineering Research and Technology, Vol. 1, pp. 1 – 4, 2012.

[4] 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, 2014.

[5] Vidushi Sharma, Sachin Rai, Anurag Dev "A Comprehensive Study of Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, 2012.

[6] Wei 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, 2012.

[7] Kamruzzaman S M, Jehad Sarkar, "A new data mining scheme using artificial neural networks", Sensors Journal, Vol. 11, No. 5, 2011.

[8] T. Karthikeyan and N. Ravikumar, A Survey on Association Rule Mining International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 1, 2014.

[9] Meenakshi Sharma, "Data Mining: A Literature Survey", International Journal of emerging research in management and technology, Vo. 3, Issue. 2, 2014.

[10] Ranno Agarwal, " Genetic algorithms in data mining", Internaitonal Journal of advanced research in computer science and software engineering, Vol. 5, Issue. 9, 2015.

[11] Kamble, Atul, "Incremental Clustering in Data Mining using Genetic Algorithm", International Journal of Computer Theory and Engineering,Vol. 2, No. 3, 2010.

[12] Shraddha Soni, "A literature review on data mining and its techniques",Indian Journal of applied research, Vol. 5, Issue. 6, 2015.

[13] Tan Jun Shan, He Wei and Qing Yan, "Application of Genetic algorithm in data mining", Proceedings of computer science conference, Vol. 2, pp. 353 – 356, 2009.

[14] Dawei, J. "The Application of Date Mining in Knowledge Management", International Conference on Management of e-Commerce and e-Government, IEEE Computer Society, pp. 7- 9, 2011.

[15] Puneet Chadha and Singh, Classifcation rules and genetic algorithm in data mining", Global journal of comptuer science and technology, software and engineering, Vol. 12, Issue. 15, 2012.

[16] Kantarcıoglu, Murat. Xi, Bowei. Clifton, Chris., "Classifier evaluation and attribute selection against active adversaries", Data Mining Knowledge Discovery Journal, Vol. 22, pp. 291–335, 2011.

[17] Diti Gupta, Abhishek Singh Chauhan, "Mining Association Rules from Infrequent Item sets: A Survey", International Journal of Innovative Research in Science", Engineering and Technology,Vol.2,Issue 10, 2013.