Performance Evaluation of Different Techniques in the Context of Data Mining- A Case of an Eye Disease

Yukti Agarwal
M.Tech. CSE Department, Amity University
Noida, Uttar Pradesh
yuktiagarwal14@gmail.com

Hari Mohan Pandey
CSE Department, ASET, Amity University
Noida, Uttar Pradesh
hari04top@yahoo.co.in

***Abstract*— This paper presents a comparative analysis of optimization techniques in the context of data mining. We have considered eye disease problems and recommendation of respective lenses. The optimization in data mining plays a fundamental role for the extraction of patterns or knowledge in minimum time. We have studied and implemented three optimization techniques such as fuzzy logic based approach, neural network based approach (perceptron based and Back-propagation based). For the experimentation purpose data from eye clinic are collected to understand the appropriate disease and recommend the type of lenses for patient. The paper also covers the analysis observations and discussions based on the obtained results.**

Keywords—Back propagation, Data mining, fuzzy logic, KDD, neural networks, optimization

I. INTRODUCTION

Data mining refers to extracting or mining the knowledge from large amount of data. The term data mining is appropriately named as 'Knowledge mining from data' or "Knowledge mining"[1].In [2], the following definition is given: Data mining is the process of exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules. Data mining is an interdisciplinary subfield of computer science which involves computational process of large data sets' patterns discovery. The goal of this advanced analysis process is to extract information from a data set and transform it into an understandable structure for further use. The methods used are at the juncture of artificial intelligence, machine learning, statistics, database systems and business intelligence. Data Mining is about solving problems by analyzing data already present in databases [3]. Data mining is also stated as essential process where intelligent methods are applied in order to extract the data patterns.

The rule extraction is the basic process of data mining. IF-THEN rules are the most common taxonomy for the rule extraction in the field of extracting knowledge from a large database. To obtain the best possible solution in the extraction, we need to model these rules in an optimized way. This paper will discuss some optimizing model of rules in the field of data mining.

In this paper a dataset of an eye clinic is taken as the example for the experimental purpose. The dataset consist of four attributes as an input and one output class attribute of a patient. The first attribute is age factor describing the age of the patient next is spectacle prescription which describes the type of spectacle the patient is using and the last is astigmatism which is a type of an eye defect. Based on these the output class will provide the type of lens recommended by the doctor.

The organization of the paper is as follows: Section-II describe the optimization techniques and their brief introduction from the previous work. Section-III presents the implementation of three optimization techniques studied using case of eye disease. We have collected the dataset from a clinic for the experiment purpose. We have done the simulation using MATLAB on the collected dataset which is shown in Section-IV. Conclusions and future work has been drawn in section-VI. Lastly, we have presented important literatures which can be used for further enhancement of the presented work.

II. OPTIMIZATION TECHNIQUES STUDIED

This paper focuses on the different techniques for rule optimization in the context of data mining. The optimization of rules can be obtained from various techniques such as fuzzy logic based, neural network based, genetic algorithm based etc. For this paper the fuzzy logic and neural network are discussed.

A. Fuzzy logic

Fuzzy logic is an approximation method to solve any problem. It contains the fuzzy sets rather than the crisp sets. The input parameters of the fuzzy logic takes only approximate values called partial

978-1-4799-4236-7/14/$31.00 ©2014 IEEE

truth not exact value [4]. The value of these parameters ranges between 0-1.These are totally different from the traditional sets which uses the values as a discrete one. Fuzzy logic is designed later to deal with the concept of partial truth, where the answer of any yes or no question is not given as quiet yes or quiet no. Here in this paper FIS (fuzzy inference system) for the given dataset is implemented in MATLAB to determine the performance of that problem with the fuzzy based approach.

B. Perceptron Network

Perceptron networks come under single layer feed forward networks and also called simple perceptrons. The perceptron network consist of three units namely, input unit, hidden unit and output unit [4]. The perceptron learning rule is given by:

$$Y=f(y_{in})=\begin{cases} 1 & \text{if } y_{in}>\theta \\ 0 & \text{if } -\theta\le y_{in}\le\theta \\ -1 & \text{if } y_{in}<-\theta \end{cases} \quad (1)$$

The weight updation of perceptron learning is as follows:

$$w_i(new)=w_i(old)+\alpha t x_i \quad (2)$$

where α=learning rate

t=target of the given pattern

θ=threshold value

The perceptron network is also implemented for the same problem using the tools of neural network in MATLAB.

C. Back propagation Network

The BPN are most important developments in the field of neural networks. It is a multilayer feed-forward network that consist of units like input units, hidden units and output units with continuous differentiable activation functions [4]. This is the network where the error is back propagated back to the hidden unit and input unit.

The net input at hidden-layer and input-layer is calculated first then net-input between hidden-layer and output-layer is calculated. Then input is mapped to the output using the different activation functions like binary sigmoidal and bipolar sigmoidal activation function. These all layers computation are inbuilt in the tool of neural network which is used here for the optimization results.

III. PROBLEM DESCRIPTION OF DATASET-AN EYE CLINIC

The problem discussed and experimented in this paper is based on the dataset of an eye clinic. Table 1 is representing the attributes and their values for a patient in context of the problem. For a patient

Table 1: Dataset of the Eye clinic

age factor	spectacle prescription	astigmatism	recommended lenses
young	myopia	yes	no lens
middle	myopia	yes	soft lens
young	myopia	yes	soft lens
aged	hyper-metrope	no	no lens
aged	myopia	yes	hard lens
young	hyper-metrope	yes	hard lens
middle	hyper-metrope	no	soft lens

IV. IMPLEMENTATION IN MATLAB

A. Fuzzy Logic Simulink:

For implementation of FIS in the MATLAB, follow the steps of the given algorithm:

S1. Open the MATLAB 7.0

S2. Write "fuzzy" in the command window and press enter.

S3. The FIS editor toolbox will get open as shown in fig. 1 then click on input and output to change the properties

S4. For each input & output

(i). Double click on the input or output and change the membership functions according to the problem as in fig. 2

S5. Now generate the rules as in fig. 3

S6. Then click on view and see the aggregated results of all rules and then at last the performance value as in fig 4.

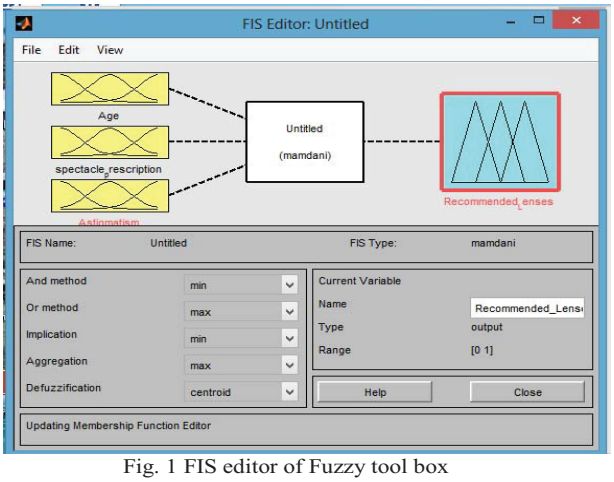

Fig. 1 FIS editor of Fuzzy tool box

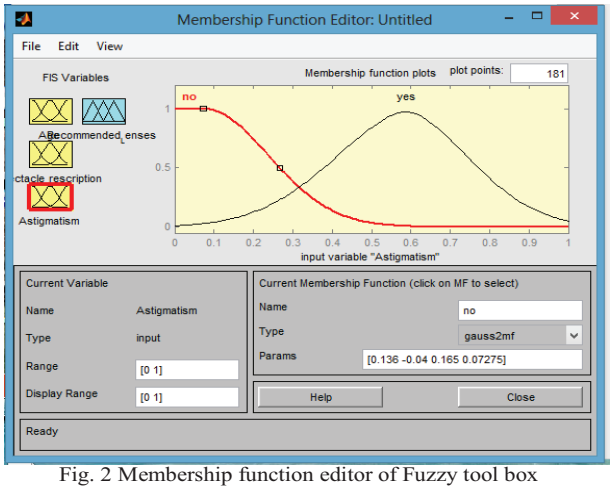

Fig. 2 Membership function editor of Fuzzy tool box

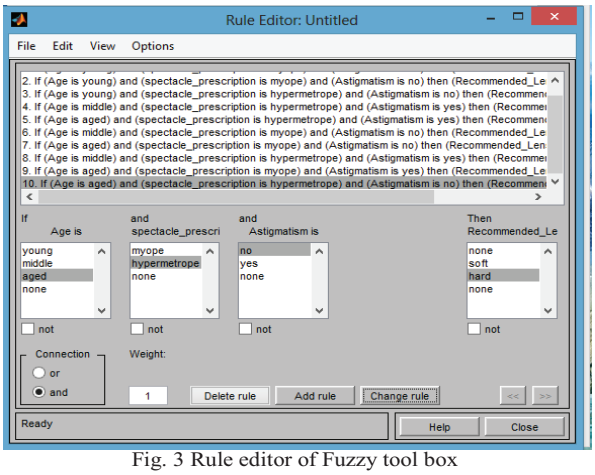

Fig. 3 Rule editor of Fuzzy tool box

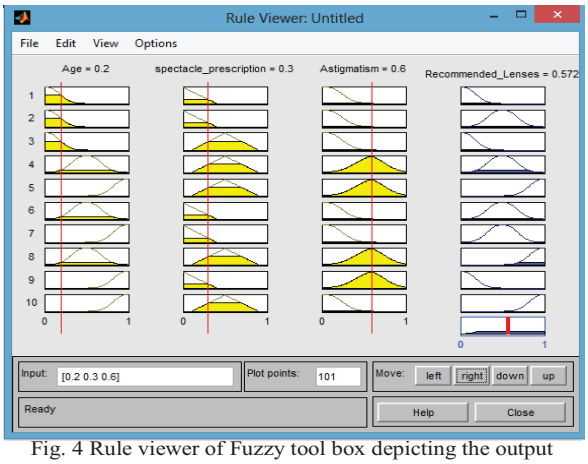

Fig. 4 Rule viewer of Fuzzy tool box depicting the output

B. Perceptron Network Simulink:

To simulate the rules using perceptron network in MATLAB we have to open neural network by writing ->nntool in command window of MATLAB and press enter. For this experiment the input and output values are taken in the binary form as shown in fig.5 & fig. 6.

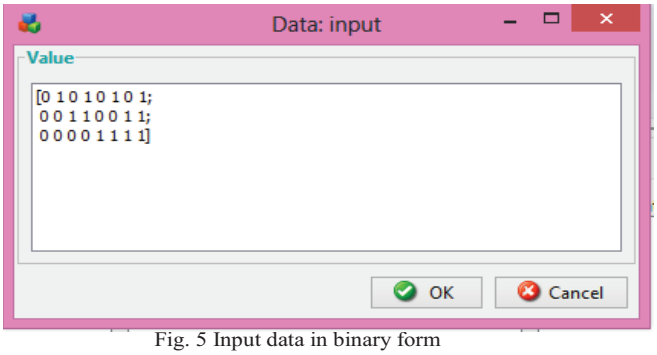

Fig. 5 Input data in binary form

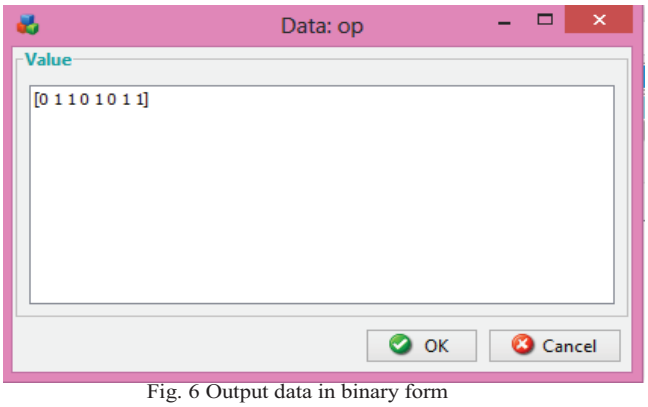

Fig. 6 Output data in binary form

After creating the codes for input and output, the network is created as shown in fig 7. Having three inputs as age factor, spectacle prescription and astigmatism.

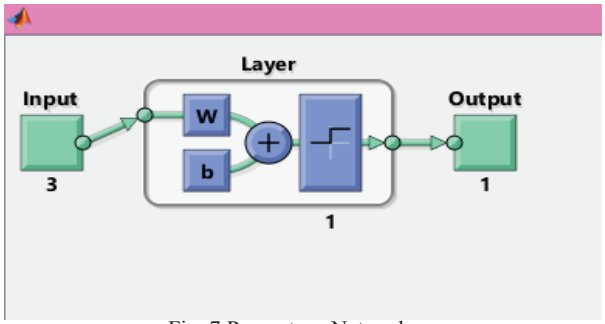

Fig. 7 Perceptron Network

Once the network is trained with the input and target values then simulated the network using same values as shown in fig. 8.

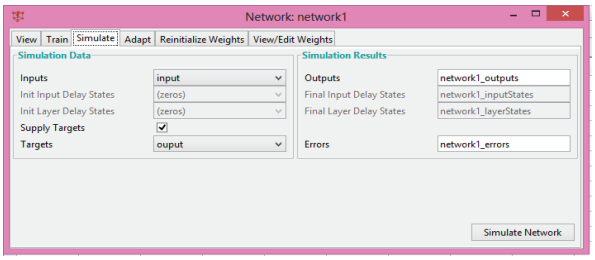

Fig. 8 Simulation Process

The resulting output after the giving input to the simulation is shown in fig. 9.

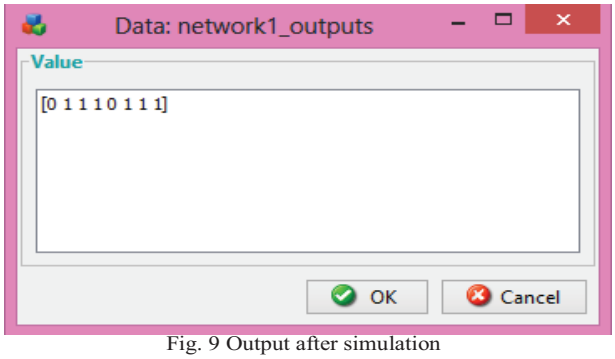

Fig. 9 Output after simulation

The error in the simulation is detected where the output after simulation is not equal to the priory given target value. The error values are shown in the fig. 10.

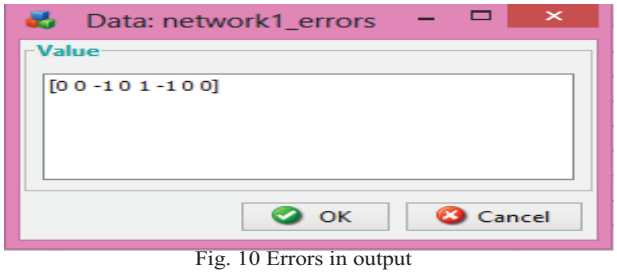

Fig. 10 Errors in output

For the above dataset the performance of the perceptron network is evaluated as 0.375 which is shown in fig. 11.

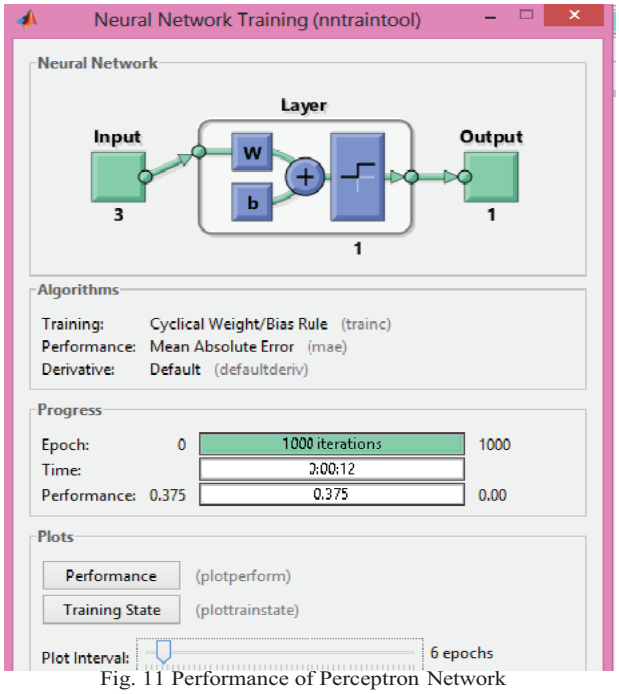

Fig. 11 Performance of Perceptron Network

C. Back-Propagation Network:

The Back-propagation network can be created by writing ->nntool in the command window and selecting the corresponding network from the dropdown list. The network for the dataset is shown in fig. 12.

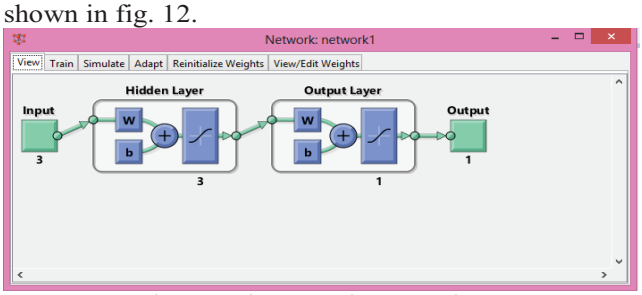

Fig. 12 Back-Propagation Network

Then the network is trained and the performance of this network obtained is 0.413 as shown in fig. 13.

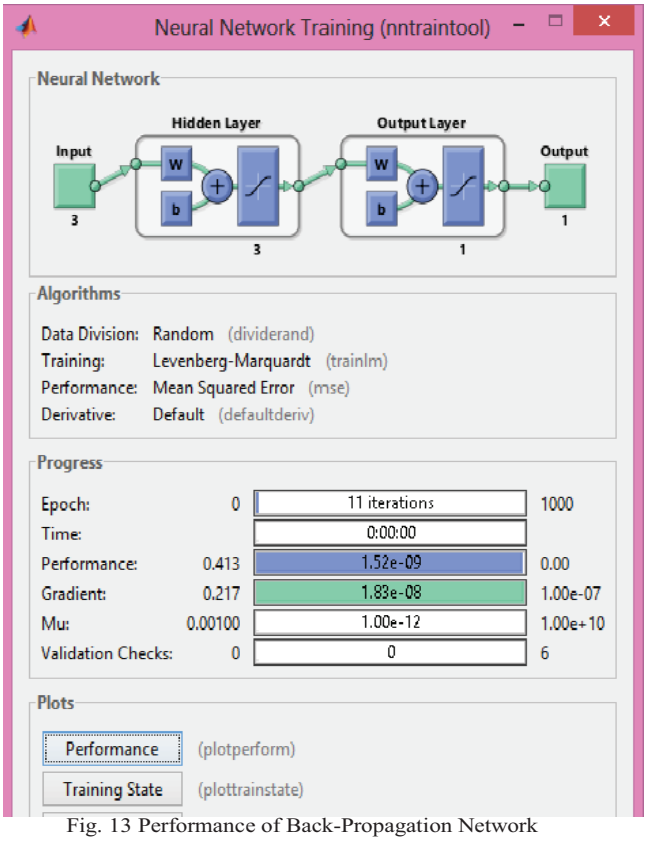

Fig. 13 Performance of Back-Propagation Network

The output after the giving input to the simulation of the network is shown in fig. 14.

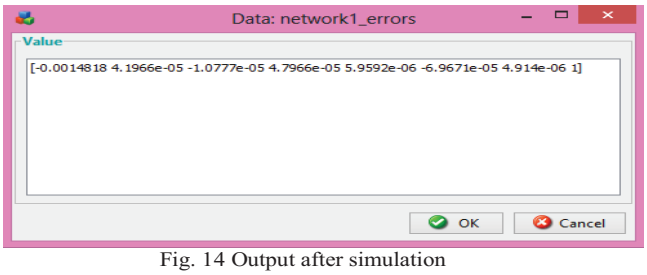

Fig. 14 Output after simulation

And the error compared to the target values are shown in fig. 15.

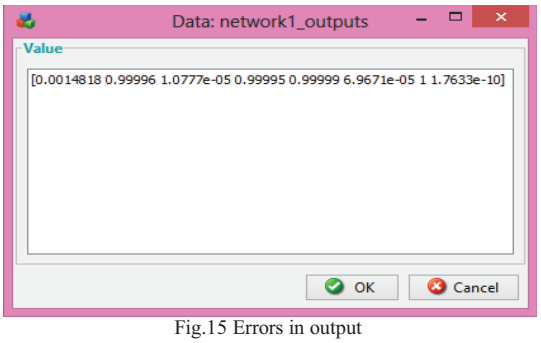

Fig.15 Errors in output

V. DISCUSSION OF THE EXPERIMENT

Table 2 presents the summary of results obtained experimentally using the collected datasets. It is evident that the performance of perceptron network is far better than Fuzzy logic and Back-propagation network. In other words we can say that perceptron network outperformed other two optimization techniques discussed in the present paper.

Table 2: Comparison of results

Technique Used	Data Set	Performance
Fuzzy Logic	Eye Clinic	0.572
Perceptron Network	Eye Clinic	0.375
Back-Propagation Network	Eye Clinic	0.413

VI. CONCLUSIONS AND FUTURE ENCHNACEMENTS

In this paper, we have studied the three optimization techniques and evaluated their performance using a dataset of eye clinic. We have conducted the experiments using MATLAB tool in order to examine the performance. The obtained results have been summarized and presented (see Table 2). It was observed that the Perceptron network outperformed other two optimization techniques considered for the experiments. During the experiments we found that Back-Propagation is showing average result whereas Fuzzy Logic is giving the worst result in the present scenario.

Although, fuzzy logic and back propagation methods used widely in several research area but for the data collected we have not achieved good results. Therefore, it would be interesting the see the impact of these two approaches on more complex cases.

VII. REFERENCES

[1] Nikita Jain,Vishal Srivastava"Data Mining Techniques: A Survey Paper" EISSN: 2319-1163 PISSN: 2321-7308, Nov-2013

[2] Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978-1-59904-252, Hershey, New York, 2007.

[3] Joseph, Zernik, "Data Mining as a Civic Duty – Online Public Prisoners Registration Systems", International Journal on Social Media: Monitoring, Measurement, Mining, vol. - 1, no.-1, pp. 84-96, September2010.

[4] S.N.Sivanandam,S.N.Deepa"Principles of Soft Computing"in Soft Computing,Wiley India Pvt. Ltd,New Delhi,pp 39-360,2008