



Data mining using neural-fuzzy tool box in Matlab

MONIKA

M.Tech Scholar, MDU Rohtak

Dr. Balkishan

Assistant Professor, DCSA Deptt. MDU Rohtak

ABSTRACT

Data mining is using a combination of neural tool box and fuzzy tool box in Matlab. It is a versatile and powerful tool; it is more capable than performing data mining. The process of data mining in MATLAB becomes a positively forward task. It is a logical extension that provides synthesis, in this type of area the creation of data mining toolbox in MATLAB is recommended. All over these tools are investigated and each tool analysis provides the recommendations for future extensions.

KEYWORDS

WSN, scalability and load balancing, cluster , graph theory

INTRODUCTION

As well as data grows, uses of the data mining tools are increasing which helps to give clean information from data sets. In a data set hidden knowledge extracting by data mining; it is a type of knowledge that is not easily obtained by queries or statistical analysis. For the classification and prediction of future events this hidden knowledge can be used.

In the development of data mining tools MATLAB has been used. For the addition, to create a toolbox necessity and feasibility is required to know about the data mining process.

MATLAB is a versatile and powerful tool and it is more capable than performing data mining. There are many tools for data mining such as Clementine, Excel and Weka but MATLAB is not yet used like as. Although more than Oracle MATLAB is more frequently chosen, and MATLAB is used with many other types of tools. This paper study ensures that the process of data mining in MATLAB becomes a positively forward task. So it is logical extensions that provide synthesis, in this type of area the creation of data mining toolbox in MATLAB is recommended.

The future opportunities of extension in this field are numerous, and this is not in the toolbox but also data mining in MATLAB.

DESIGN CONSIDERATIONS

For design, it is necessary to find the capabilities to carry out different types of data mining process and recommend extensions to synthesis of these tools. It is great work to check the potential of MATLAB in data mining field. For this study we need to understand data mining in a proper pattern and this is as-

- All details about data mining in a broad context.
- The data mining process in every stage.

Data Mining Introduction: For the analysis of the data mining process, we will need a brief look. Their names are like as-

- Supervised learning
- Unsupervised learning
- Hybrid learning

Supervised Learning:- In the field of data mining process most popular technique is Induction Based Supervised. In this type of technique, it involves model creation and training data set testing and the model of application is classified in new data. Supervised learning is classified in three categories such as:

- Classification (This deals with current behavior and categorical data)
- Estimation (This deal in current behavior but the data have continuous numerical data)
- Prediction (This deals with types of data and its future behavior)

Unsupervised learning:- Unsupervised learning is totally opposite to the supervised learning model, during clustering performance no dependent variable exists. In the unsupervised clustering idea is to find the concept structures within data and the questions regarding the formulation of data set, these were never be asked before. The clustering demonstrates that there is any useful information exists in the data set and supervised learning model help us for best inputs. (E.g. in the search of underline)

Hybrid Learning:- The data mining realistic approach of the potential is realized when combination of supervised and unsupervised learning is implemented is called as Hybrid learning. The aim of this study is to discover the potential of MATLAB during the synthesis of different data mining tools.

Different algorithms do not work same in all areas, so that it gives us a sense to implement different varieties of algorithms, especially when we work on a big project. The evaluation tool for supervised learning is unsupervised clustering. Beginners must be selected carefully when carrying out supervised learning.

III THE PROCESS OF DATA MINING: We will classify the data mining process in 4 different phases.

- Decision (Here we will take decision on our data it will be carried out or not)
- Data preparation (Here all the analysis of data is reading)
- Model building (Here a predict model is carried out for building from our work)
- Interpretation (The graphical representations of the results)

In the flow chart Figure no. 2.1 shows the process. Each stage of the process discussed below briefly, it will be supervised, unsupervised or a combination. The preparation for our tool assessment is serving as.

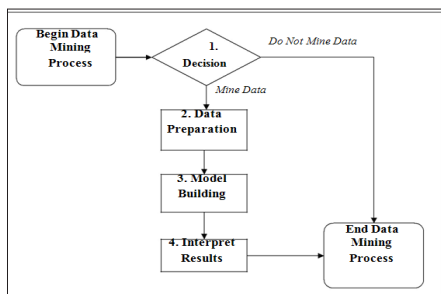


Figure 1: Data Mining Process Phases

Decision Phase:- In data mining decision is the first stage where we decide to go with given data analysis or not. For all the stages it is important and also difficult. We decide whether we spent our study time or not and a given data set discover other resource. This is the stage for human where he can ask and reply the answer to find the defined problem. It helps us to taking of the first step.

Data Preparation Phase:- For building model tool we will interpreted our data which will be in proper data format that we want. All these things are decided here. The Data preparation is mostly very large topic. According to this study we will need to understand the importance of our data preparation for the study. We will also understand the investigated facilities of the tools.

In data preparation examples, it includes:

- Outliers search,
- Continuous data discrete,
- Normalization.

Model Building Phase:- The data mining process have a core process which is known as model building means main task of the model. It is place which give results, those are verified. This study has limited scope for supervised learner. Because for the known cases our created model is worked e.g. they are using verified results which are never being presented in the model from the information of data. All these are done in a stage which is known as testing, validation and training stage.

Interpretation Phase:- In data mining process interpretation is the final and last stage. This stage indicates as, "The human element provides the analysis of result that data mining project can be failure and success ultimately". Due to the decision phase, using automated means the results interpretation can be assisted.

METHODOLOGY: The study is large so it breaks down in two different phases.

1st, we will analyze and assessment the package. This is the place where we claims the tool for validation and suggest the possible outcomes to the tools.

2nd, we will synthesize the package. Where we used combinational tools for obtaining final result to implement the process.

Tool Assessment:-The extended documentation of the tool is essential in Tool Assessment. For obtaining these types of results we will need some experiments on tools. From all these types of case studies this case is the best example. In first stage of the tool assessment the critical evaluation of claims. In data mining process the terms of evaluation is claims that have been pertained. When we get valid claims with suggestions then we will be more improved to fulfill its purpose.

Synthesis:- For a specific purpose the process of building or designing a fresh (new) idea is defined as Synthesis. This can be done by taking the parts together in a proper pattern. Which we want to do in this study this is closest, the chosen tools of the methodology carry out data mining by MATLAB is

true potential of this work. In the data mining tool by MATLAB there is no means of synthesis exists.

MATLAB is chosen for data mining which is broad field must exist for available tools in synthesis. In the first stage process synthesis of the tools is decide. According to this study the discussion will be focus on the synthesis of Math Works Fuzzy Clustering, Data Analysis Toolbox, Neural Network Toolbox and ARMADA.

Essentially, when tools work together for each other complement is highlighted and the synthesis which is suggested then implemented.

FINAL ANALYSIS

In the final, the whole process needs evaluation. Finally this is crucial a holistic view to obtain by the data mining synthesis process in MATLAB. In the final stage results are obtained in order to determine what we have been found and what we do in the new process. In many ways, it can be showed as "tool".

As showed in Figure 2 below, it is a circular process. For obtaining a complete overview of MATLAB all facts make it clearer. The present available tools of data mining are accessed here and synthesized again. In many important ways MATLAB will gain from this. The main finding in MATLAB is able to process the data mining. This can be says on the basis of the evidenced through the extension of this work.

Figure shown below shows broad methodology as:

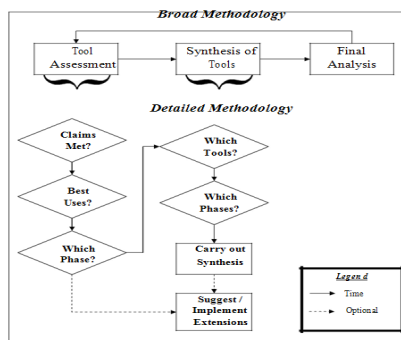


Figure 2: Data Mining Synthesis Tools in Broad and Detailed Methodology

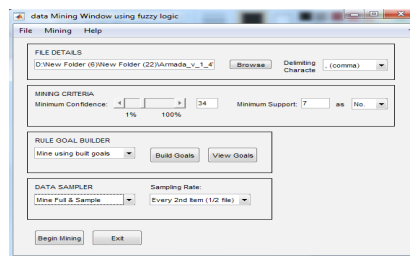


Figure 3: Using Fuzzy logic Data Mining Window

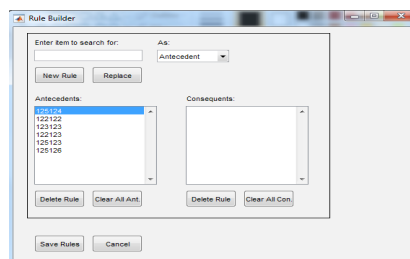


Figure 4: The rule builder index

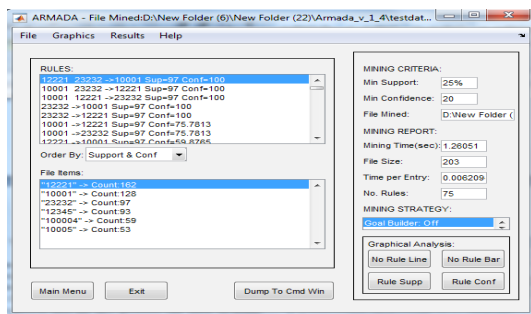


Figure 5: rule analyzer index

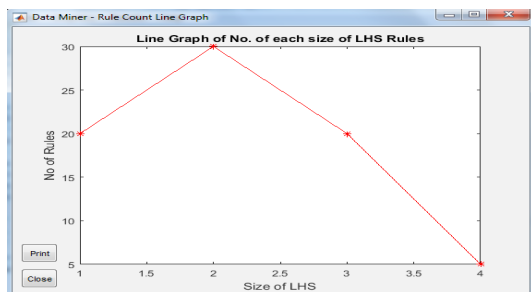


Figure 6: Number of rule line graph

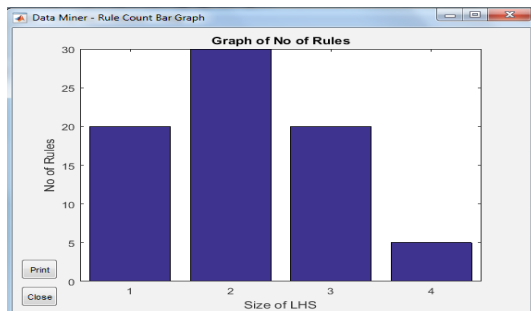


Figure 7: Number of rules graph

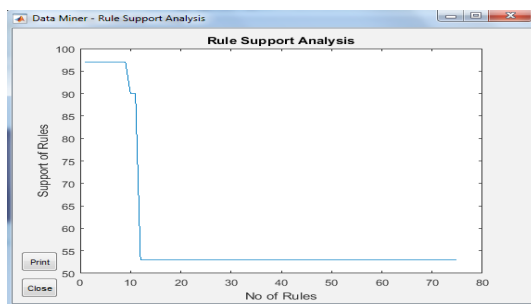


Figure 8: Number of rules with rules support

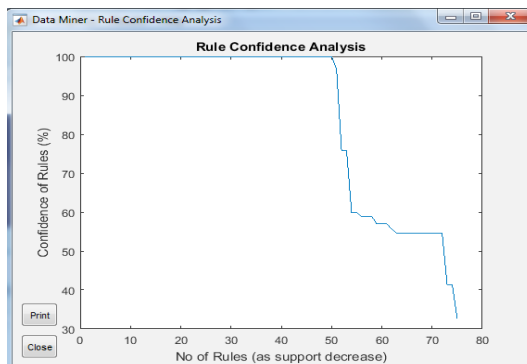


Figure 9: Number of rules with rule confidence

RESULT & CONCLUSION

In this work the aim is to obtain synthesis and analysis of data mining process from MATLAB toolbox. In this study, a methodology which can be used in future and have many possible extending in creation for developing a toolbox of data mining. There are many important findings that are summarized and discussed below:

- The phases of the data mining process is outlined and refined (as shown in Figure 1)
- For synthesizing MATLAB data mining tools established a methodology (which shown in Figure 2) and also find confirmation with experiment of its usefulness
- Skeleton scripts of neural network provide, and here it also simplifying neural network creations
- Limitations of neural network finds and these are as:
 - Results by nature "Black Box"
 - Facilities of post-and Pre-processing
- By using Fuzzy Clustering and Data Analysis Tool we will verify the claims:
 - Provide examples
 - Study separate cases
- During Fuzzy Clustering tool's documentation it will be highlighted the poor English which is used
- Many limitations of ARMADA is Highlighted:
 - In entirely numerical data sets it is very difficult to obtained the data
 - Poor handling when data set is large
 - Limited facilities of data handling and data preparation

REFERENCES

• Data Mining; Adriaans, P. and Zantige,(1997) • Addison Wesley(1988), England, pp between 39-127,(1997) • Data Analysis and fuzzy clustering Toolbox in Matlab; Balasko, B., Veszpremi University, Hungary, (2006) • Data Analysis and fuzzy clustering Toolbox in Matlab:- Feil, Veszpremi University Hungary, (2006) • Mathematical Programming; Burton, M., pp 1-10, University of Rhodes, South Africa, (2006) • Neural Networks; Burton , , University of Rhodes, South Africa, pp 101— 115, (2006) • Neural Networks; AKKAYA, K. AND YOUNIS, (2005) • Mathematical Programming Languages; Dvinnell, (1998) • Principles of Data Mining; Hand, Mannila, H. and Smyth, MIT Press, Cambridge, Massachusetts, pp 1—24, (2001) • < http://www. kd nuggets.com/polls/>; KDnuggets, KDnuggets Past Polls, (2006) • Cancer diagnosis via linear programming; Mangasarian, O.L., and Wolberg, Number 5, pp 1,18, (1990) • linear programming; W.H., , Siam News, Volume 23,(1990) • Bayes Net Toolbox for MATLAB :- Murphy,(2006) < http:// bnt.sourceforge.net >