IJCTA, 9(26) 2016, pp. 97-102
© International Science Press

A Survey on Data Mining and Soft Computing Tools

R. Sathyaraj[1] and S. Prabu[2]

ABSTRACT

Soft computing mainly used to find the solution which is an approximate and inexact solution. In this paper, we discussed various soft computing and data mining tools used in different applications to achieve better predictions. We initiated with techniques in soft computing that helps to the human as computational intelligence and the tools applied in various areas. This paper gives the idea about the tools and merits of using those data mining and soft computing techniques.

Keywords: Data mining, soft computing, classification, clustering.

1. INTRODUCTION

In new age eras, every one accepted data mining techniques and soft computing techniques such as Neuro-Computing, Evolutionary Computing, Fuzzy logic, and Genetic Computing for various studies. In the last few years these techniques individually as well as in the hybrid form solved varieties of problems in various range of areas. It provides training in the computational intelligence field and improves the research and develops demanding applications in wide-ranging domain [1].

Data mining is a computer science field; it mines the data or intelligent information from the large database with some computational process. Main properties of data mining are, discovering patterns automatically, creating information by focus on large databases.

Soft computing is an approximate reasoning, search & optimization, neural networks, fuzzy logic and evolutionary algorithms. It is differed from hard computing, basically hard computing focused on crisp data and numerical analysis. Soft computing mainly focussed on approximate values and dynamic situation or real time systems. It also deals with noisy data, parallel computing and approximate answers.

Soft computing applied in various applications like handwriting recognition, image processing and data compression, automotive systems and manufacturing, soft computing to architecture and agriculture, decision-support systems, neuro fuzzy systems, fuzzy logic control, machine learning applications, speech and vision recognition systems, and in most of the predictions [2].

2. DATA MINING AND SOFT COMPUTING TOOLS:

Studies showed the classification, clustering and prediction on different data sets using seven computing tools.

A. WEKA (Waikato Environment for Knowledge Analysis)

Weka performs data mining process with collection of machine learning algorithms like Naïve bayes, Random forest, etc. we can also call Weka through java, otherwise directly using GUI editor we can apply

[1] School of Computer Science and Engineering, VIT University, Vellore, Tamilnadu, *E-mail: sathyaraj@vit.ac.in*; *sprabu@vit.ac.in*[2]

the algorithms for various dataset. It is possible to use big data in Weka tool. It also has classification, clustering, regression, association rules, data pre-processing and visualization [3].

Using this Weka GUI visualization, it is possible to achieve better prediction model and data analysis. We can import the data in different formats like csv, arff file format, etc.

Classification can be processed with the help of classify panel. It gives accuracy resulting in prediction. Also it classifies the resultant data as false positive, true positive, precision, recall and ROC curves.

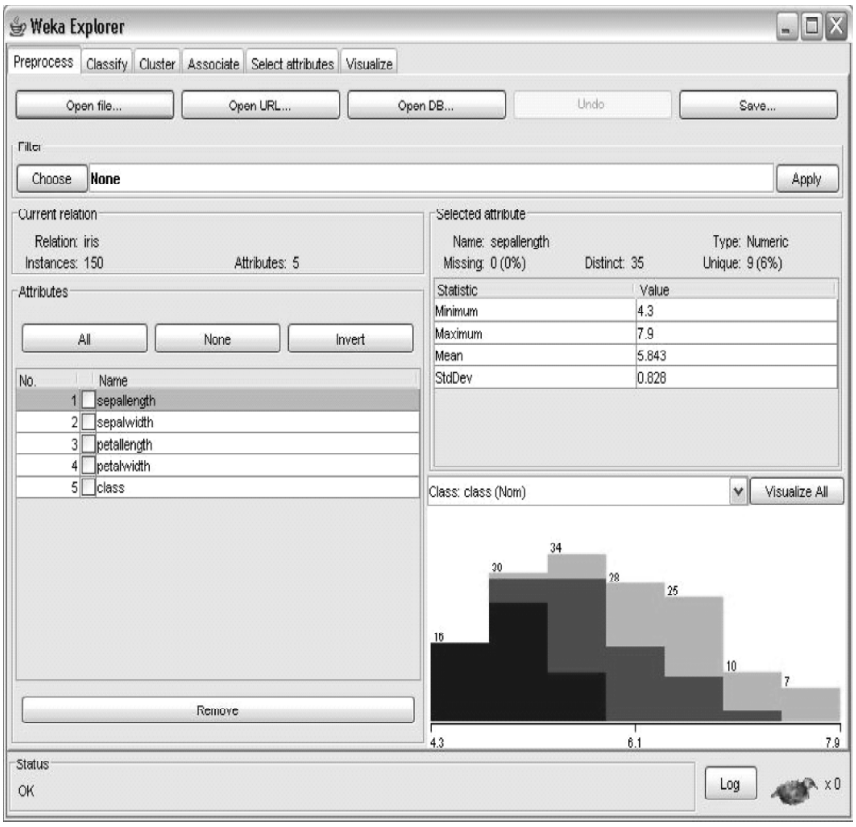

Figure 1: WEKA Explorer

B. Rapid Miner

Rapid miner is a tool for data mining applications. It provides best prediction for various problems virtually using machine learning algorithms. It has plus R and python as it extension facility. Rapid miner is a self-service predictive analytics tool.

Rapid miner provides reusability with collective building blocks, process and templates. It also provides a single platform for text and data mining, business analysis, machine learning and predictive analysis.

Rapid miner footsteps on business application and commercial application, also it focused education, data mining with machine learning procedures including transformation and visualization, research and optimization. Basic Rapid miner is an open source model licensed under AGP.

Rapid miner marketplace improves the functionality of rapid miner using additional plugins. This marketplace supports to create data analysis algorithms. Finally it is the tool that performs best.

C. FuzzyTech

It is the fuzzy logic and neuro-fuzzy based tool and it is the leading tool. Fuzzy logic is the logic provides approximate results rather than crisp values [2]. Fuzzy dealing with vagueness and this is simply fast and adaptive to any environment.

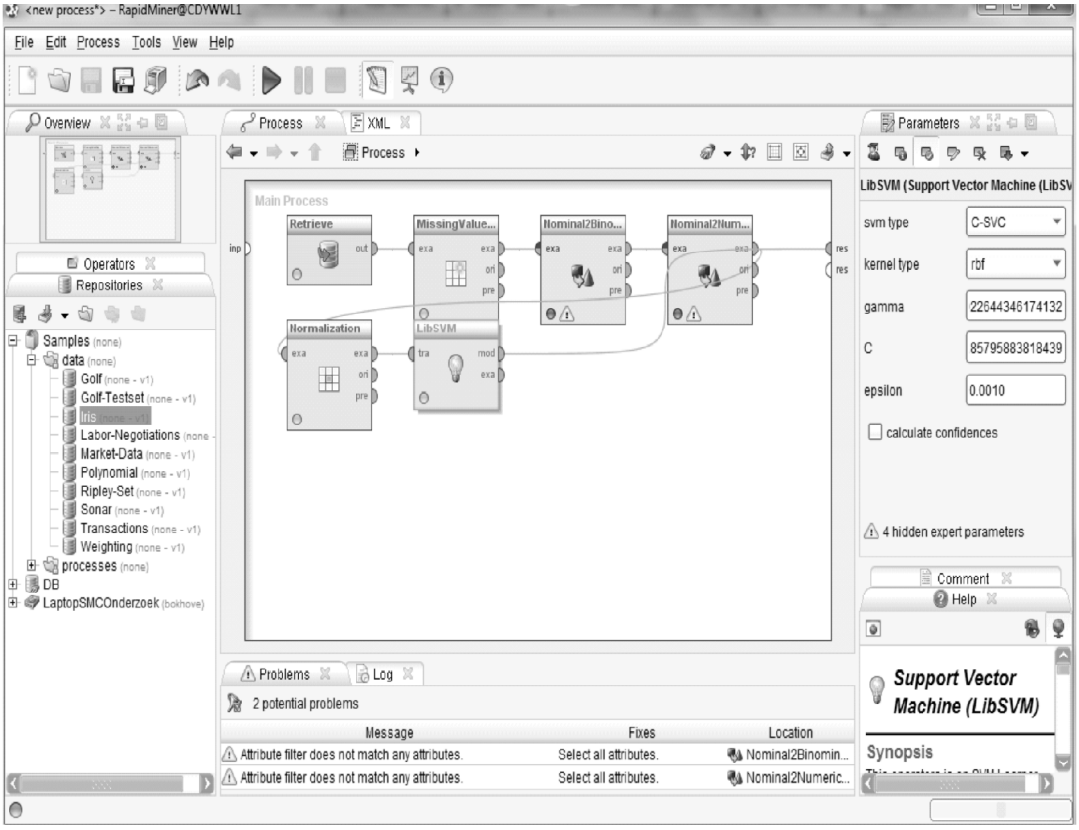

Figure 2: Rapid miner GUI

In this tool we have comfortable text editor and rules of the fuzzy systems defined by means of table-or-matrix editor. At any time it is possible to switch from one mode to another. FuzzyTech's Linguistic Variable Editors enable to simply "draw" various types of membership functions with the mouse.

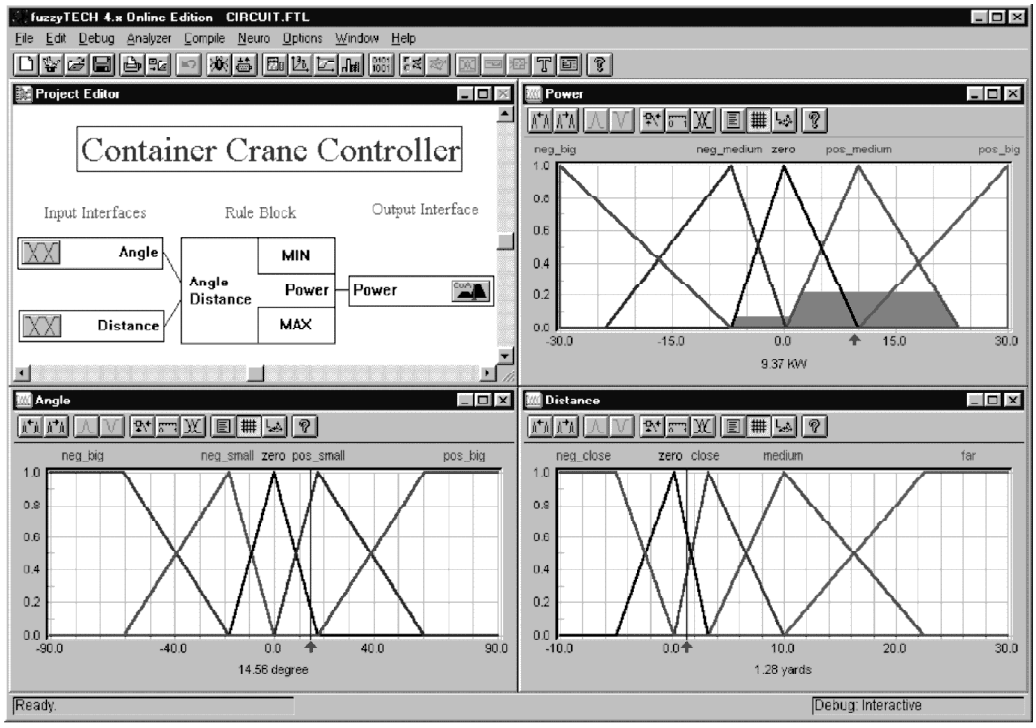

Figure 3: FuzzyTech tool

D. R studio

Main advantage of R studio, it is free and open source and with the statistical language, it's generally considered to be very easy to code. Actually, commercial and open source data analysis and visualization software increasingly integrating with R. Some de-merits are command prompt, lack of GUI is intimidating, poor parallelization, difficulty handling big data.

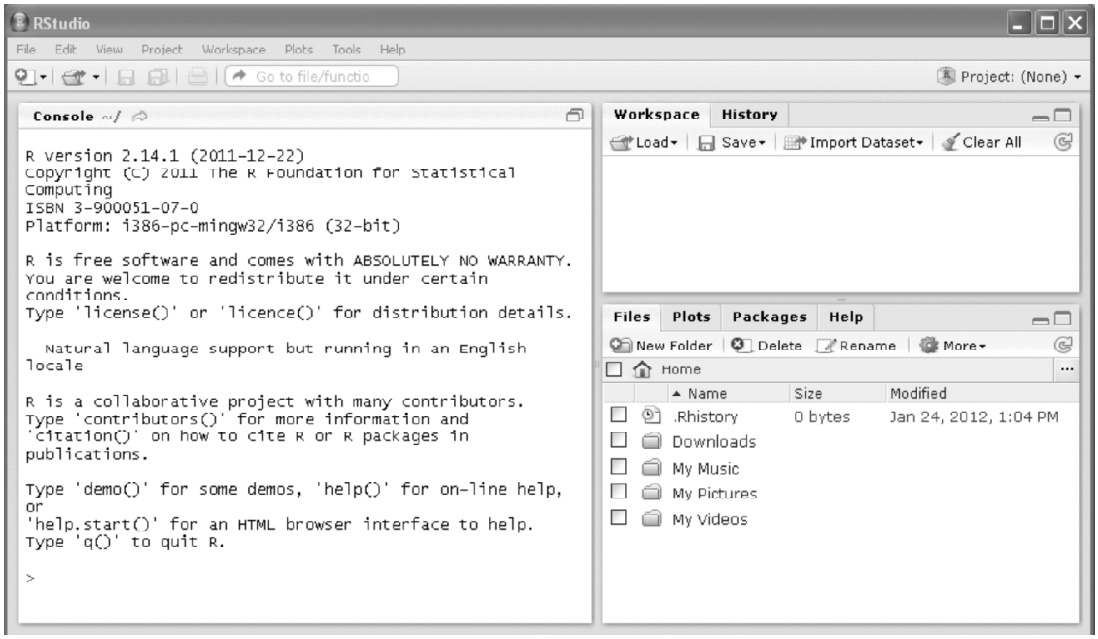

Figure 4: Rstudio tool

E. Orange

Orange is an open source data mining tool. It has very strong data visualization capabilities and it allows the users to use drag and drop modules. Orange Canvas GUI connects them to assess and test many machine learning procedures on the dataset. Setting up Orange tool is very easy and getting knowledgeable with its GUI components. With suitable dataset we explore some visualization widgets included with Orange.

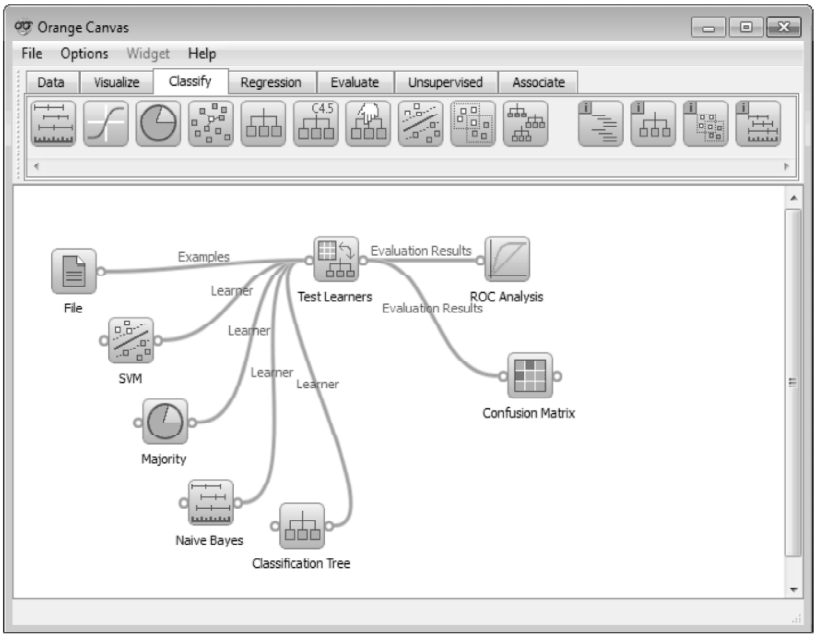

Figure 5: Orange tool GUI

F. FANN

The Fast Artificial Neural Network library (FANN) is a free open source neural network library. Neural network solve many problems which is based on human neurons [4].

FANN implements multilayer artificial neural networks and supports both entire and lightly connected networks. Cross-platform execution in both fixed and floating point is supported. It contains a framework for effective handling of training data sets.

FANN have many advantages like easy to use, adaptable, well recognized, and fast [5].

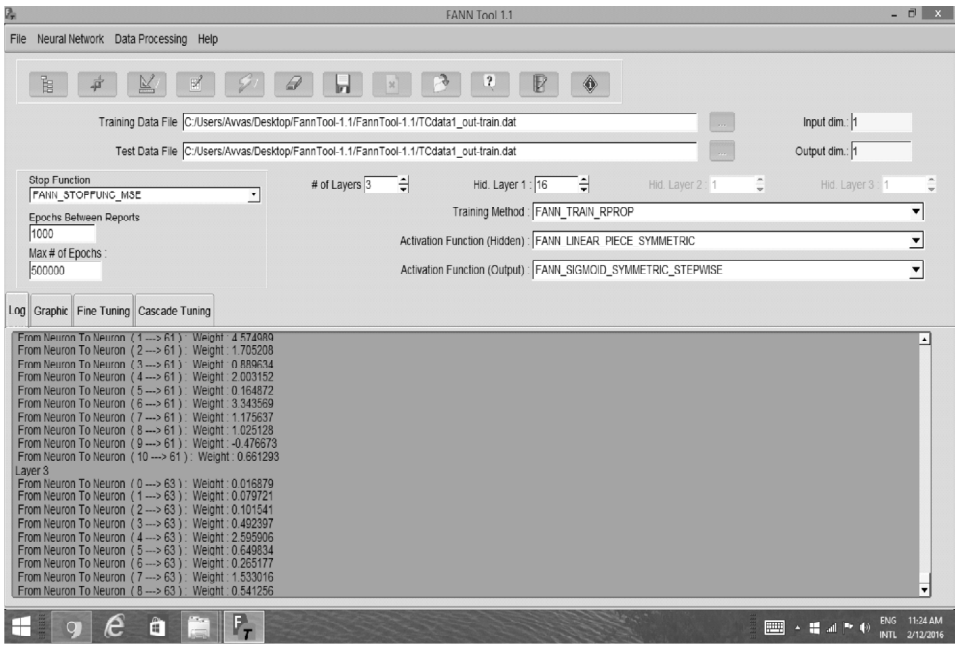

Figure 6: FANN tool GUI

G. MATLAB

MATLAB is the high-level language interactive tool, developed by MathWorks. It is used by lots of researchers, engineers and scientists and also it provides visualize ideas across disciplines.

MATLAB includes datamining, control systems soft computing, signal processing, image processing, and data analysis. It provides tools for applications with custom GUI. Mostly for the research MATLAB is the main tool for researcher.

3. CONCLUSIONS

In this paper, the main motive of this work is to explore the tools and techniques used in data mining and soft computing field. This paper expresses the basic idea about the tools and pros and cons about the tools. Researcher can better perform in their research area using these tools and these new methodologies are performing vital role in new development.

REFERENCES

[1] Prasad MC, Florence L, Arya A., "A Study on Software Metrics based Software Defect Prediction using Data Mining and Machine Learning Techniques", *International Journal of Database Theory and Application*,**8(3)**, 179-190, 2015

[2] Zadeh, Lotfi A., "Soft computing and fuzzy logic", *IEEE software*, **11(6)**, 48, 1994.

[3] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH., "The WEKA data mining software: an update", *ACM SIGKDD explorations newsletter*, **11(1)**, 10-18, 2009.

[4] Khoshgoftaar TM, Allen EB, Hudepohl JP, Aud SJ., "Application of neural networks to software quality modeling of a very large telecommunications system", *IEEE Transactions on neural networks*, **8(4)**, 902-909, 1997.

[5] Nissen, Steffen., "Implementation of a fast artificial neural network library (fann)", *Report, Department of Computer Science University of Copenhagen (DIKU)*, **31**, 29, 2003.