An Overview of Software Tools in Data Mining Systems

R. Dharmarajan[1] T.Buvitha[2]
[1]Assistant Professor [2]M.Phil Research Scholar
[1,2]Department of Computer Science & Engineering
[1,2]Thanthai Hans Roever College, Perambalur-621212, India

Abstract— Data mining is a rapidly growing field, whose development is driven by strong research interests as well as urgent practical, social, and economical needs. The growth and relevance of data mining algorithms involves the use of powerful software tools. As the number of available tools prolongs to grow, the option of the most suitable tool becomes increasingly difficult. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. This paper attempts to give overview of data mining tools and its congregation.

Key words: Data Mining Tools, WEKA, Knime, Tanagra, Rapidminer

I. INTRODUCTION

Brain tumor is naturally serious and life intimidating because of its character in the limited space of the intracranial cavity (space formed inside the skull). Most Research in developed countries proves that the numbers of people who have brain tumors were died due to the fact of inaccurate detection. Magnetic resonance images are a very useful tool to detect the tumor growth in brain. MRI is directed into intracranial cavity produces a complete image of brain. This image is visually examined by the physician for detection and diagnosis of brain tumor. However this method of detection resists the accurate determination of stage and size of tumor. To avoid that, the computer aided analysis for brain tumor detection is proposed. The image processing techniques allows the detection of tumor tissue with accuracy and reproducibility comparable to manual detection. In addition, it also reduces the time for analysis. At the end of the process the tumor is extracted from the MR image and its exact position and the shape also determined. The following figure embraces the fundamental steps in image processing system.

II. DATA MINING TOOLS

Different types of data mining tools are available in the marketplace, each with their own strengths and weaknesses. The description of the data mining tools used in this comparative study.

A. WEKA Tool:

WEKA, formally called Waikato Environment for Knowledge Learning, is a computer program that was developed at the University of Waikato in New Zealand for the purpose of identifying information from raw data gathered from agricultural domains. WEKA supports many different standard data mining tasks such as data preprocessing, classification, clustering, regression, visualization and feature selection. The basic premise of the application is to utilize a computer application that can be trained to perform machine learning capabilities and derive useful information in the form of trends and patterns [3].

WEKA is an open source application that is freely available under the GNU general public license agreement. Originally written in C the WEKA application has been completely rewritten in Java and is compatible with almost every computing platform. It is user friendly with a graphical interface that allows for quick set up and operation. WEKA operates on the predication that the user data is available as a flat file or relation, this means that each data object is described by a fixed number of attributes that usually are of a specific type, normal alpha-numeric or numeric values. The WEKA application allows novice users a tool to identify hidden information from database and file systems with simple to use options and visual interfaces. The following figure 1 presents the WEKA GUI chooser [3].

Fig. 1: WEKA GUI chooser

B. Knime:

KNIME (Konstanz Information Miner) is an open API workflow based data mining tool that provides easy accessibility to new nodes to be added into the workflow. It provides its user with the GUI which aid with the simplification of workflow generation by the user. It also provides with features to modify a particular node accordingly and execution of partial data flow [5].

C. Tanagra:

This extension of SIPINA provides the users an easy to use interface for the analysis of either real or artificial data. It allows the researchers to easily add their own data mining research methodology or any newly identified data mining processing technique and also supports by providing them with architecture and a means to compare their methodology performances. It provides the beginners or naives with a platform where they can carry out their experimental procedures [4].

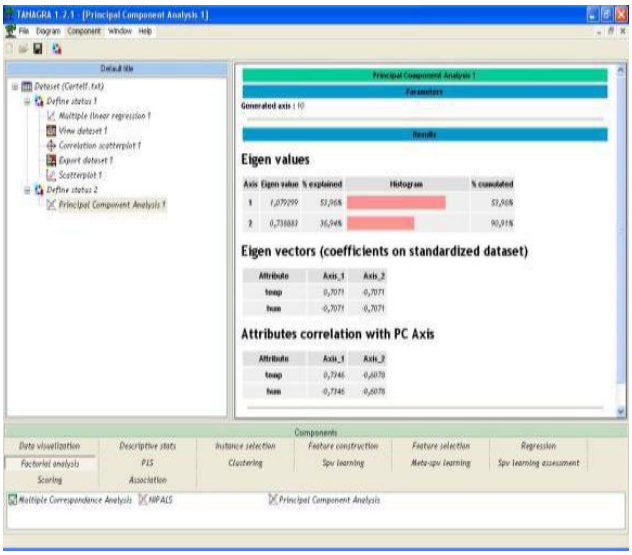

Fig. 2: Tanagra GUI

D. Rapidminer Tool:

RapidMiner[6], formerly YALE (Yet Another Learning Environment), is an environment for providing data mining and machine learning procedures including: data loading and transformation (ETL), data preprocessing and visualization, modelling, evaluation, and deployment. The data mining processes can be made up of arbitrarily nestable operators, described in XML files and created in RapidMiner's graphical user interface (GUI). RapidMiner is written in the Java programming language. It also integrates learning schemes and attribute evaluators of the Weka machine learning environment and statistical modelling schemes of the R-Project. RapidMiner can be used for text mining, multimedia mining, feature engineering, data stream mining and tracking drifting concepts, development of ensemble methods, and distributed data mining. RapidMiner is found in the: electronics industry, energy industry, automobile industry, commerce, aviation, telecommunications, banking and insurance, production, IT industry, market research, pharmaceutical industry and other fields. The following figure 3 shows the GUI for RapidMiner.

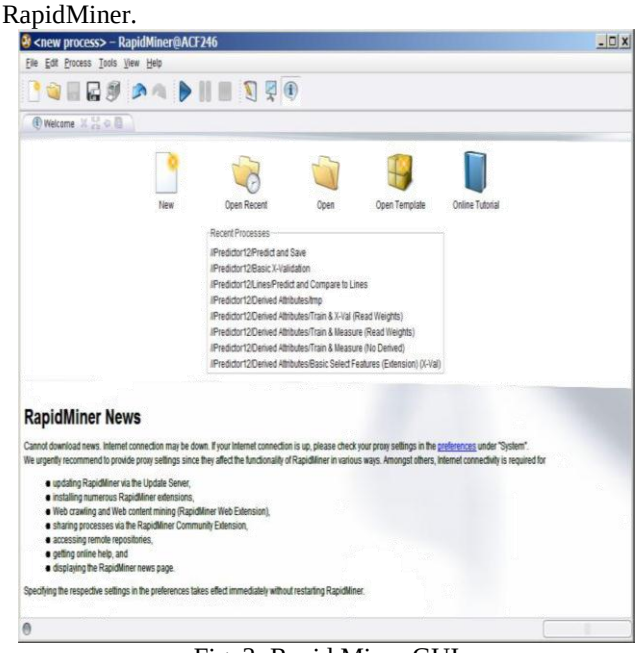

Fig. 3: Rapid Miner GUI

E. DBminer Tool:

DBMiner, a data mining system for interactive mining of multiple-level knowledge in large relational databases, has been developed based on our years-of-research. The system implements a wide spectrum of data mining functions, including generalization, characterization, discrimination, association, classification, and prediction. By incorporation DBminer performs interactive data mining at multiple concept levels on any user-specified set of data in a database using an SQL-like Data Mining Query Language, DMQL, or a graphical user interface. Users may interactively set and adjust various thresholds, control a data mining process, perform roll-up or drill-down at multiple concept levels, and generate different forms of outputs, including generalized relations, generalized feature tables, multiple forms of generalized rules, visual presentation of rules, charts, curves, etc. The figure 4 shows GUI for DBminer tool [7].

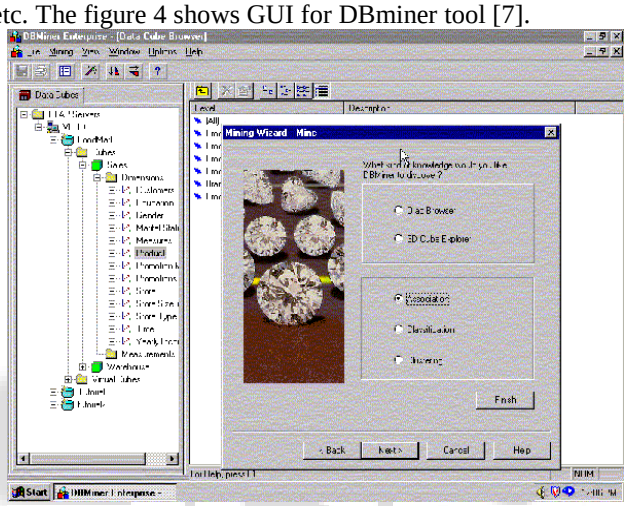

Fig. 4: DBminer GUI

III. MATCHING OF DATA MINING TOOLS INTO ALGORITHMS

The following table matching the different types algorithms and the number of recent tools in data mining system.

Tool Name	Decision Tree	Linear/Statistical	Multi-layer Perceptron	Nearest Neighbor	Radial Basis Functions	Bayes	Rule Induction	Polynomial Networks	Generalized Linear Model	Time Series	Sequential Discovery	K Means	Association Rules
CART	Yes												
MARS		Yes											
SEE5	Yes						Yes						
DIGITAL LOGIC	Yes						Yes						
GRIT BOT													
SAS	Yes	Yes	Yes		Yes				Yes	Yes		Yes	Yes
MAGNUM OPUS													Yes
SPSS		Yes							Yes				
LERS													Yes
WEKA	Yes	Yes	Yes			Yes	Yes						

Table 1: Matching Tools and algorithm in data ming

IV. ADVANTAGES AND DISADVANTAGES

The following table describes the advantages and disadvantages of different types algorithms in data mining system

Tool Name	Advantages	Disadvantages
CART	Depth of tree options	Difficult file input/output, limited visualization
MARS	Multivariate regression	Limited algorithms
SEE5	Depth of tree options	Few data options, limited visualization
DIGITAL LOGIC	Can be working with imprecise, incomplete and inconsistent data	Limited algorithms
GRIT BOT	Strong anomalies finding	Limited algorithms
SAS	Depth of algorithms, visual interface, reasonable graphic output	Hard to use, need to master SAS programming language
MAGNUM OPUS	Fast speed (linear computation time), good flexibility, can handle big data set	Limited usage
SPSS	Depth of algorithms, widely used, powerful functionalities	Hard to use
LERS	Easy to use	Limited algorithms, no visualization tools
WEKA	Can be working on any platform	Low speed

Table 2: Advantages and Disadvantages

V. CONCLUSION

The growing problem of inaccessibility of information due to the amount of data derived from the area is known as data mining. Software tools are needed to make data mining applications efficiently. The paper presented the specific details along with description of various data mining tools enlisting the area of specialization.

REFERENCES

[1] Han, J., Kamber, M., Jian P., Data Mining Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers, 2011.
[2] Abdullah H. Wahbeh, Qasem A. Al-Radaideh, Mohammed N. Al-Kabi, and Emad M. Al-Shawakfa "A Comparison Study between Data Mining Tools over some Classification Methods" IJACSA,Special Issue on Artificial Intelligence.
[3] The WEKA data mining software: An update, Mark Hall, Eibe Frank, G. Holmes, B. Pfahringer, P. Reutemann, IH Witten, ACM SIGKDD Explorations, Newsletter, Pages 10-18, volume 11 issue 1, june 2009.
[4] "A Study of Data Mining Tools in Knowledge Discovery Process", Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K.Ratnam.
[5] KNIME-Available-at: http://www.knime.org/download-desktop
[6] http://rapid-i.com
[7] DBMiner: A System for Data Mining in Relational Databases and Data Warehouses, Data Mining Research Group, Intelligent Database Systems Research Laboratory School of Computing Science, Simon Fraser University, British Columbia, Canada, http: db.cs.sfu.ca DBMiner

