Use Classification Algorithms to Build a Classifier Model

Vivek Kumar Sharma[1], Sudhir Kumar Verma[2], Dr. Navin Kumar Tyagi[3]

[1]M.Tech Scholar, Department of Computer Science and Engineering, MIT, Bulandshahr

[2]AssistantProfessor, Department of Computer Science and Engineering, MIT, Bulandshahr

[3]Professor, Department of Computer Science and Engineering, MIT, Bulandshahr

Abstract- In today is data mining landscape, there are numerous tools available, both open source and commercial, that offer solutions for solving data mining problems. These tools provide various strategies, such as decision trees, neural networks, and lazy classifiers, to address classification problems. Each strategy involves a multitude of parameters that users can customize, including epochs, learning rate, momentum, and more, in the case of neural network classifiers.

I. INTRODUCTION

In the actual world, we are inundated with vast quantities of information, which continues to grow exponentially. This data often exists in its raw form, concealing a wealth of valuable information [1]. The process of data mining involves taking out this hidden, implicit, and earlier unidentified data-derived information. It involves creating computer programs that automatically search databases for consistencies or trends. In addition to numerous tools accessible, both open-source and commercial, to tackled at a mining problem.

One widely used open-source tool in this domain is WEKA (Waikato Environment for Knowledge Analysis) [3]. WEKA offers a variety of strategies, including decision trees, neural networks, and lazy classifiers [4]. Each strategy allows users to customize many parameters. For instance, given the scenario of a neural network classifier, users need to mention values for parameters such as epochs, learning rate, and momentum [3].

Although tools provide default parameter settings, it has been observed that classifier act, in points of accuracy, frequently enhanced by conducting experiments with various parameter assess. Consequently, for inexperienced users, it becomes challenging to determine the appropriate values, and their only recourse is to engage in time-consuming trial and error.
Furthermore, the classification process must be repeated every time a new data set is provided.

II. DATA MINING

The process of data mining of interaction with the knowledge base or user. The required trends are offered to the client, and which can be kept as unique information within the database. Data (Information) mining is important step, because it discovers unseen, previously unknown patterns for evaluation.

The process of learning involves a series of important steps that are iterative in nature:

- Integrated data (several data bases could be mixed).
- Selecting information (information which is necessary for the purpose of analysis are selected out of the database).
- Transformation of data (the conversion of data must be converted into the right forms for mining.)
- Mining of data (an important procedure in which intelligent approaches are used to find various data patterns)
- Pattern Evaluation (It is verification process).
- Knowledge Presentation (Illustration and presentation methods are employed to show the extracted information to the user).

III. WEKA GUI

- The University of Waikato in New Zealand is where
WEKA was created [3], [4].

According to figure1.7, Six tabs are present in the Explorer window. It supports basic operations as follows:
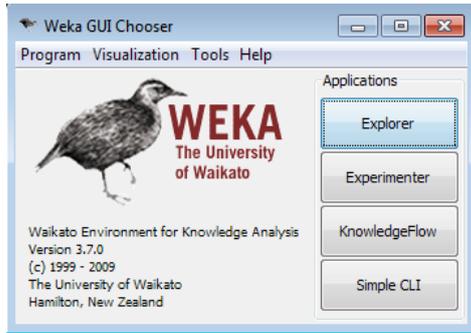
Preprocess: This task allows selecting the data set and modifying it in different way.

Classify: It allows training of learning schemes that execute classification and assess them.

Cluster: It allows learning of clusters for the dataset.

Associate: It performs learning of evaluate the data and assign rules to it.

Choose attributes: The best feature in the dataset is selection.

Visualize (Imagine): It is used to view various two-dimensional plots of data.

IV. MODELLING AND IMPLEMENTATION

The important consideration made in this thesis is to use existing standard classifiers and have developed innovative algorithm with different functionality. In this chapter, we will study different design aspects that are useful in successful implementation of system.
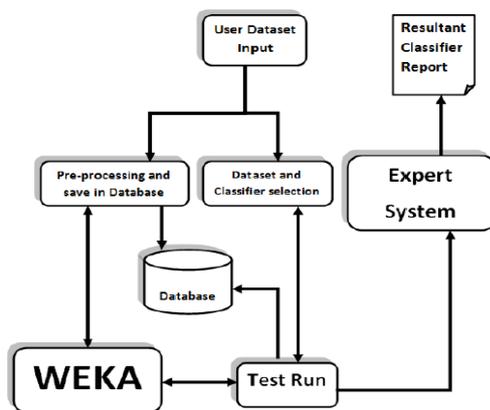
Above diagram show the Architectural Diagram.

The Novice User will input the collective information in the form of .arff (attribute related record configuration) file. This data will be given to the WEKA (Waikato Environment for Knowledge Analysis) along with some dynamic parameter.

Expert System will then, apply the classifiers on the dataset which are already defined in WEKA, the classifiers are applied one by one on the dataset and the statistics of the respective classifiers are saved in the MS Access database file in separate relation.

When statistics of all classifiers are successfully stored in the data base then the next and very important thing is done i.e. selection of the best classifier, which is done by comparing the statistics of all classifiers, and the best classifier is shown to the user along with its statistics. The user can also see the statistics of other classifiers.

V. ALGORITHM

Input- Information file in .arff (Attribute Relation File Format) format.
Output- The classifier's name and precision with parameter values to be used.
Process: It is separated into two halves, A and B.

A) For determining:

1. To acquire the best accuracy, read the database file that contains the parameter values we are looking for.
2. Perform "P*Q" times. For instance, Q is the number of algorithms, and P is the how many times the same algorithm is used is performed with various criteria values.
3. Execute algo. L to Q.
4. Give the criteria values $L=X$
   Criteria 2=Y
   Criteria 3=M
......
......
Criteria B=b s
5. Result for analysis
6. View the output file that was produced.
   Retrieve the perfect value from the source input.
7. Save algo, Criteria L. ,Criteria B values and perfect result to the database.
8. Move to step3 till "P*Q" till performing action.

B) Suggestion criteria values and classifier algorithm to user:

1. To obtain the best result, study the information file on which we wish to determine the classifier method and criteria settings.
2. In the result file, look up the database name.
3. If result is not found, show the result "No record found" and repeat step 2 else show the criteria

values, perfect value and name of the classifier from the output file.

4. Stop.

VI. RESULT

Data Preprocessing Window in which User can provide training dataset/relation information. User can give input using New File command or through Open DB command or may use Open URL option.
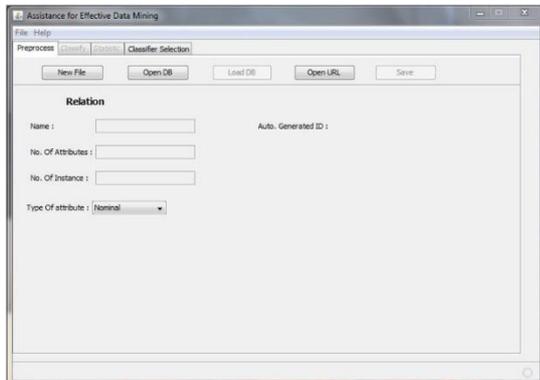
New file: User can give database in arff (Attribute Relation File Format).

Open DB: User can provide MS Access.

Load DB: This option can be used only when user provides MS Access database so that to load database in preprocess panel and convert it into attribute relation file format.

Open URL: Used to accept database from URL Text Fields are as follows:

Name: Relation name.

No. of Attributes: Total type of information terms of relation.

Total Instances: Total types of occurrence in relation.

Type attribute: Which types of attributes are present in relation? (Nominal/Numeric/Mixed)

Auto. Generated ID: Each relation is assigned a unique id number for further use and it is displayed.

VII. CONCLUSION

We have presented new approach to store parameter values of previously mined training dataset. User Assistance performs analysis of training data and provide best classification algorithm with its parameter values. User can use these values in future on another test data which is relatively same as training data. It can save time of user by providing details of classifier based on past experience. Classification Algorithm plays very important role in User Assistance as they reduces time requirement for analysis. It can help user in decision making in any field of work. Diagnosis, business, marketing, banking sectors can make use of this User Assistance in their work.

It can be used in marketing and business field where decision making for day to day activity is very important.

REFERENCE

[1]. K.Cios,W. Pedrycz, R. Swiniarski, L. Kurgan, "Data Mining: A Knowledge Discovery Approach", ISBN 978-0-387-33333-5, Springer, 2007.

[2]. Mehmed Kantardzic, "Data Mining: Concepts, Models, Methods, and Algorithms", ISBN 0471228524, Wiley-IEEE Press, 2002.

[3]. Remco R. Bouckaert, Eibe Frank, Mark Hall,"WEKA Manual for Version 3-6-8", University of Waikato, Hamilton, New Zealand, August 13, 2012.

[4]. Ian Witten, Eibe Frank and Mark A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques", 3rd Edition, ISBN 978-0-12-374856-0, Morgan Kaufmann, 2005.