Prediction and Analysis of students Behavior using BARC Algorithm

M.Sindhuja
Assistant Professor, Computer Science & Engineering
Sri Shanmugha College of Engg. & Tech.
Salem, India
sindhumega@gmail.com

Dr.S.Rajalakshmi
Associate Professor, Computer Science & Engineering
Jay Shriram Group of Institutions
Tirupur, India
mrajislm@gmail.com

S.M.Nandagopal
Associate Professor, Computer Science & Engineering
Easa College of Engineering
Coimbatore, India

***Abstract*— Educational Data mining is a recent trends where data mining methods are experimented for the improvement of student performance in academics. The work describes the mining of higher education students' related attributes such as behavior, attitude and relationship. The data were collected from a higher education institution in terms of the mentioned attributes. The proposed work explored Behavior Attitude Relationship Clustering (BARC) Algorithm, which showed the improvement in students' performance in terms of predicting good behavior, average attitude and good relationship with faculty members and Tutors. The Hierarchical clusters were grouped with related similarities and analysis was experimented using WEKA tool. The resulted analysis describes the input parameters were found optimal.**

Keywords- Behaviour, Attitude, Relationship, Hierarchical Clusteingr, Students' Performance and Prediction.

I. INTRODUCTION

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [1]. There are increasing research interests in using data mining techniques in educational filed. This new emerging field, EDM, concerns with developing methods that discover knowledge from data originating from educational environments.

Educational data mining techniques often differ from traditional data mining techniques, in explicitly exploiting the multiple levels of meaningful hierarchy in educational data.

1.1 EDM METHODS

Romero and Ventura [2007] categorize work in educational data mining into the following categories:

a) Statistics and visualization

b) Web mining

(i) Clustering, classification, and outlier detection

(ii) Association rule mining and sequential pattern mining

(iii) Text mining

Baker, which classifies work in educational data mining as follows:

a) Prediction

(i)Classification

(ii)Regression

(iii)Density estimation

b) Clustering

c) Relationship mining

(i)Association rule mining

(ii)Correlation mining

(iii)Sequential pattern mining

(iv)Causal data mining

d) Distillation of data for human judgment

e) Discovery with models

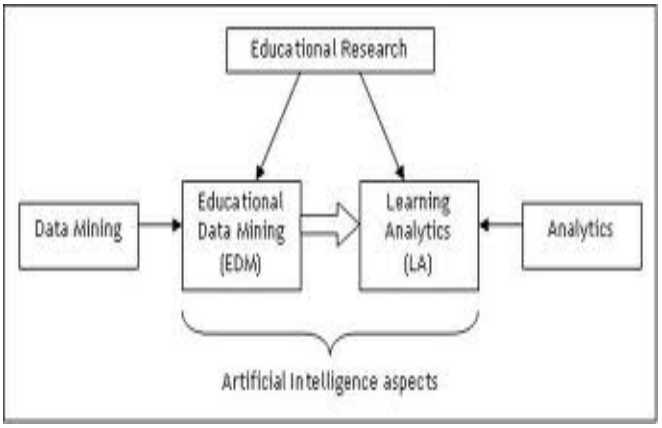

Figure 1. Educational Data miming

1.2 DATA MINING AND TECHNIQUES

Data Mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing knowledge sharing to make proactive, knowledge-driven decisions.

The most commonly used techniques in data mining are:

1. Artificial Neural Networks: this is a nonlinear predictive model that learns through training and resembles biological neural networks in structure.

2. Decision trees: tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.

3. Genetic Algorithms: They are optimization techniques that use process such as genetics combination, mutation, and natural selection in a design based on concepts of evolution. 4. Rule Induction: the extraction of useful if-then rules from data based on statistical significance.

5. Regression Methods: this tries to identify the best linear pattern in order to predict the value of one characteristic we are studying in relation to another.

Data mining does various types of tasks like,

Classification is a task of finding a function that maps records into one of several discrete classes.

Clustering is a task of identifying groups of records that are similar between themselves but different from the rest of the data.

II. LITERATURE REVIEW

Zaïane (2001) suggests an application for data mining, using it to study on-line courses. This article proposes and evangelizes EDM's usefulness, and in this fashion was highly influential to the formation of our community.

Zaïane (2002) and Tang and McCalla (2005) center around how educational data mining methods (specifically association rules,and clustering to support collaborative filtering) can support the development of more sensitive and effective e-learning systems.

Tang and McCalla report an instantiation of such a system, which integrates clustering and collaborative filtering to recommend content to students.

Baker, Corbett and Koedinger (2004) gives a case study on how educational data mining methods can be used to open new research areas, in this case the scientific study of gaming the system.

Merceron and Yacef (2003) and Romero et al. (2003), present tools that can be used to support educational data mining.

Beck and Woolf (2000) shows how educational data mining prediction methods can be used to develop student models. They use a variety of variables to predict whether a student will make a correct answer. This work has inspired a great deal of later educational data mining work – student modeling is a key theme in modern educational data mining, and the paradigm of testing EDM models' ability to predict future correctness.

III. DATA COLLECTION

The students' data were collected as a dataset from a Department of an institution as shown in Figure 2. To predict the attributes such as behavior, attitude and student staff relationship interaction, the similar set of data were predicted and clustered under a tutor. Where, tutor is a teaching staff and they are responsible for the set of student under them. As it is related with the students' cognitive intelligence, using tutoring system the data were collected for experiment.

name	sex	year	student behaviour	stud attitude	knowldege distribution	SSR	STR	TSR
ram	m	3	1	0	-1	1	1	1
rama	f	3	0	1	1	-1	0	1
raji	f	3	1	0	-1	1	1	1
shoba	f	3	1	0	-1	1	1	1
sooraj	m	3	1	0	-1	1	1	1
anil	m	3	1	0	-1	1	1	-1
nair	m	3	1	0	-1	1	1	-1
radha	f	3	1	0	0	1	0	1
priya	f	3	-1	0	-1	1	1	1
sowmiya	f	3	0	0	-1	1	-1	1
chitra	f	3	-1	0	-1	1	-1	1
mani	m	3	-1	0	-1	1	0	-1
raj	m	3	-1	0	1	1	0	-1
sabari	m	3	0	-1	1	1	-1	-1
lalitha	f	3	1	0	-1	1	1	-1
punitha	f	3	1	0	1	1	1	1
mala	f	3	1	0	1	1	1	1

Figure 2. Student data set with attributes

IV. BARC ALGORITHM

The basic idea of clustering is to define the similarity between the distance, the distance that represents the data between the data to measure the similarity of the size of the data are classified, until all the data gathering is completed. The proposed work uses hierarchical clustering analysis using the parameters as behavior, attitude and relationship. Hierarchical clustering creates a hierarchy of clusters which may be represented in a tree structure called a dendrogram. Clustering process to construct a spanning tree, which contains the class hierarchy information, and the similarity between all classes and class, it generates a hierarchical nested class, and can be used for any feature type, high accuracy. But because each merger, the global compare all the similarity between the class and choose from the smallest two classes, so run slower, not suitable for large-scale data sets.

BARC algorithm describes students' Behavior, Attitude and Relationship with tutor and Staff. The parameters correlation can be experimented using matrix structure. The attribute behavior is termed as (α), Attitude as (β) and relationship as (γ).

The relation between the attributes is as shown below.

D = Ci (α+β+γ)……..Ci..n (α+β+γ), Where α,β,γ be 0,1,-1 …….(1)

From the equation 1, the matrix of each clusters are derived.

BARC ALGORITHM

(1) Choose D as a dataset;

(2) Data Collection, C = (α,β,γ), of Ci…n (α,β,γ)i….n, where

D is to identify the clusters of optimal similarities relates to the attributes behavior, attitude and relationship.

(3) IF C=1 AND ||D||>1 THEN

(4) Assign that student in that clusters as a optimal behavior, attitude and relationship;

(5) Randomly select a center point data, that exhibits the student performance when α,β,γ = 1 or -1.

(6) Repeat to group a cluster until α,β,γ applies to 1 or -1.

(7) If student α,β,γ = 0, then avoid the student from the cluster.

V. EXPERIMENT USING WEKA

The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. It is freely available software. It is portable & platform independent because it is fully implemented in the Java programming language and thus runs on almost any modern computing platform. Weka has several standard data mining tasks, data preprocessing, clustering, classification, association, visualization, and feature selection. The WEKA GUI chooser launches the WEKA's graphical environment which has six buttons: Simple CLI, Explorer, Experimenter, Knowledge Flow, ARFFViewer, & Log.

5.1 Preprocessor

The Preprocess panel imports the data from a database, a CSV file, ARFF etc., and preprocesses this data using filtering algorithm which can be used to transform the data from one format to other e.g. numeric attributes into discrete ones. It is also possible to delete instances and attributes according to specific criteria on the preprocess screen. It is also possible to view the graph for particular attribute.

Figure 3. depicts the attributes processing which shows the Students attributes level.

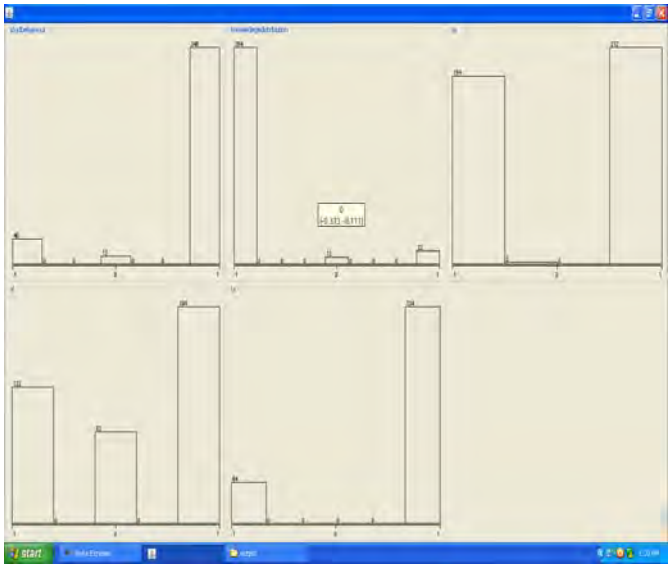

Figure 3. Preprocessor Output

5.2 CLUSTERING

The Cluster panel is used to access the clustering techniques in Weka, e.g., the simple k-means, EM, DBScan, XMeans algorithm. Sometimes it is necessary to ignore some attribute while using the clustering algorithm, so it is possible with Ignore Attribute button.

The proposed work experimented using DBScan clustering. Figure 4. depicts 8 cluster from the input dataset, where 6 input were found noisy.

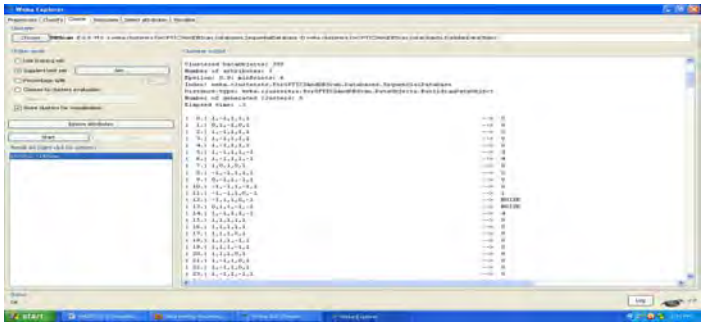

Figure 4. DBScan Result of Cluster

5.3 HOT SPOT TECHNIQUE

The Associate panel gives access to association rule e.g. Apriori, Predictive, Apriori algorithm. Once the appropriate parameter for association rule is chosen then result list allows the result set to viewed or saved.

Hot Spot learns a set of rules (displayed in a tree-like structure) that maximize/minimize target variables. With a nominal target, one might want to look for segments of the data where there is a high probability of a minority value occurring (given the constraint of a minimum support). For a numeric target, one might be interested in finding segments where this is higher on average than in the whole data set.

From the input dataset, the association to the target attributes was found increased. Figure 5. depicts the association with the input and the target attribute. We fix the normal branching factor as 2. It shows that there is some improvement in behavior, attitude and also in student faculty relationship.

In BRICHcluster, the maximum cluster is generated by the given input dataset is 9 from X0 to X9. It is relatively optimal, and it is predicted that ratio of the dataset is found strengthened.

This will help to analyze the noisy factor, which is low and can be eliminated by adopting the tutoring system in higher education combining of all the year students to a tutor.

Figure 6. depicts the BRICHCluster output and Figure 7. depicts visualization of the patterns.

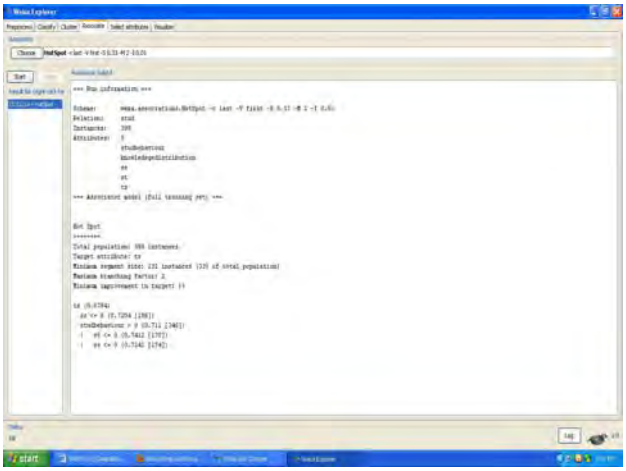

Figure 5. Association of Attributes

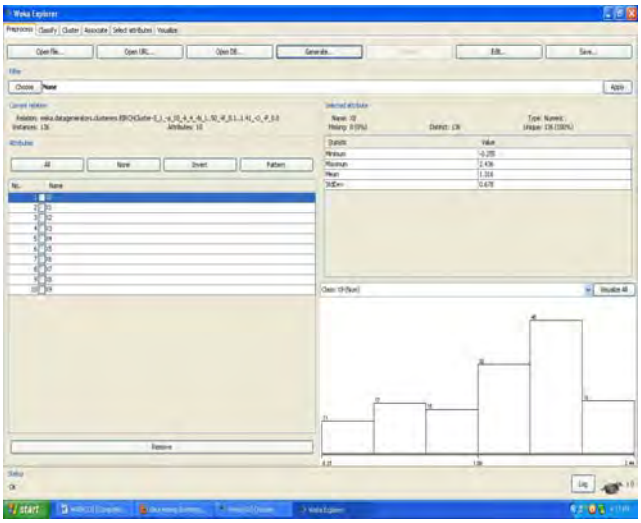

Figure 6. BRICHClusters

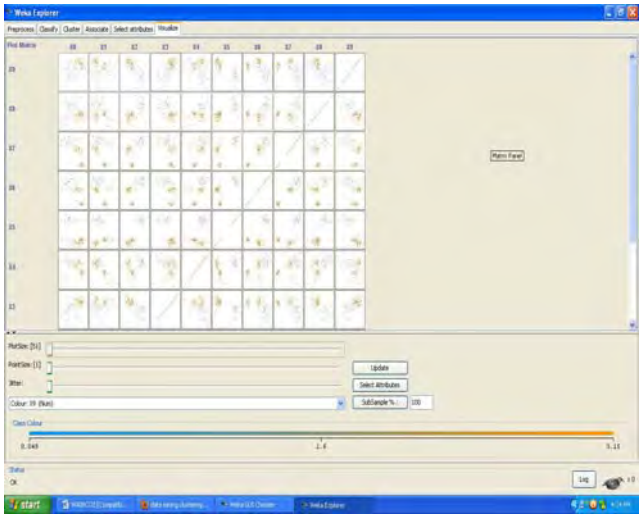

Figure 7. BRICHClusters Visual Patterns

VI. CONCLUSION

The research work explored prediction and analysis of student behavior, attitude and relationship of student and staffmembers. BARC Algorithm dicussed about the prediction and elimination of cluster patterns which is not similar to optimal attribute relations.Weka tool preprocessed the student dataset which showed percentile of the student of higher education had good behavior, average attitude and optimal relationship with staff and tutor. The maximum clusters found to be generated using BRICHClusters and DBScan analysis. It exhibits the system to the improvement of teaching learning process towards the academic improvement of students in higher education.

REFERENCES

[1] Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning, with application to clustering with side-information. In: Advances in Neural Information Processing Systems, vol. 15, pp. 505–512. MIT Press, Cambridge (2003)

[2] Kosala, R., Blockeel, H.: Web mining research: A survey. SIDKDD Explorations (July 2000)

[3] Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief, U.S. Department of Education Office of Educational Technology, October 2012.

[4] E. Ferneley, R. Helms, “Editorial of the Special Issue on Social Networking,” Journal of Information Technology, Vol.25, No.2, pp. 107- 108, 2010.

[5] Educational Data mining as a Trend of Data Mining in Educational System, Sachin 7. R. Barahate, International Conference & Workshop on Recent Trends in Technology, (TCET) 2012 Proceedings published in International Journal of Computer, pp. 11-16.

[6] Barnes T., “The q-matrix method: Mining student response data for knowledge”, In Proceedings of the AAAI-2005 Workshop on Educational Data Mining, 2005.

[7] Brijesh Kumar Baradwaj, Saurabh Pal, “Mining Educational Data to Analyze Students’ Performance”, In International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, pp. 63-69, 2011.

[8] Castro F., Vellido A., Nebot A., & Minguillon J., "Detecting atypical student behavior on an e-learning system", In I Simposio Nacional de Tecnologas de la Informacin y las Comunicaciones en la Educacin, Granada, pp. 153–160, 2005.
[9] Romero C., & Ventura S., "Educational data mining: A survey from 1995 to 2005", Expert Systems with Applications, 33(1), pp. 135–146, 2007.
[10] Scott J., "Social network analysis: A handbook", (2nd ed.). Newberry Park, CA: Sage, 2000.
[11] Kurgan. L and Musilek. P (2006), A survey of knowledge discovery and data mining process models, Knowledge Engineering Review, 21(1), pp: 1–24.
[12] Mizan b.Hitam, Sabariah bt Hj Mahat and Rajasegaran .K (2002)",The tacit Knowledge Dimension for Knowledge Management in Higher Education Organizations".
[13] Petko Valtchev, Rokia Missaoui, and Robert Godin (2004)" Formal Concept Analysis for Knowledge Discovery and Data Mining: The New Challenges".
[14] Rowley, J. (2000), Is higher education ready for knowledge management? The International Journal of Educational Management, 14(7), pp. 325-333.