International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

Improving Students' Learning Process by Clustering Of Students' Data

Bhavana A. Khivsara[1], Nisha S. Nahata[2], Madhavi A. Bhattad[3] , Shraddha S. Savant[4], Rahul Kumar[5]

[1]*Department of computer Engineering, SNJB's KBJ COE, Chandwad, bhavana.khivsara@gmail.com*
[2]*Department of computer Engineering, SNJB's KBJ COE, Chandwa, nishanahata28@gmail.com*
[3]*Department of computer Engineering, SNJB's KBJ COE, Chandwad,mbhattad1@gmail.com*
[4]*Department of computer Engineering, SNJB's KBJ COE, Chandwad,shraddhasavant10@gmail.com*
[5]*Department of computer Engineering, SNJB's KBJ COE, Chandwad, rahulkumar9797@gmail.com*

Abstract- Few years ago there were just a handful of educational institutes that could provide few courses to small amount of students. Today the institutes are not only providing courses to thousands of students but also produce enormous amounts of data of students from their systems but it remains unused. The proposed system uses that data to asses students. The System show how data mining algorithms can help discovering academically relevant knowledge contained in databases obtained from educational institute Web access Log. These findings can be used both to help teachers with handling their class, understand their students' learning and reflect on their teaching and to support student reflection and provide proactive feedback to Students.

Keywords-Data Mining, Educational System, Web Access Log, Educational Data Mining.

I. INTRODUCTION

Now a day's huge amount of data is being generated. Data mining is the process of determining useful data from large amount of data stored in database or database warehouse [1]. Now days, the educational institutes are facing several problems such as identifying students requirement, personalization of training and predicting quality of student and identifying their interest. Manual data analysis of such a huge data creates bottleneck. In order to overcome these issues Educational data mining (EDM) provides a set of techniques to improve quality of students as well as automate tutors' work [2]. Data mining software allow user to analyze data from different perspective, categorized it and summarized the relationship, identified during mining process. This study analyze students' web access log by using clustering and shows students learning behavior and performance. So it will become easy for educators to keep watch on students' and guide them accordingly. Teachers' can guide or advice students' according to necessity. More specifically, it will improve the monitoring work of educators, and will automate some of the educators' work.

The further paper is organized as, the section II: gives broad survey on EDM, section III: cover mathematical foundation of system, in section IV: the proposed system architecture is discussed, section V: shows dataset and experimental results and in section VI: conclusion is given.

II. LITERATURE SURVEY

In this section the study of existing system, their working, advantages and limitations are covered. According to a recent survey conducted by Campus Computing (campuscomputing.net) and WCET (wcet.info), almost 88% of the surveyed institutions reported having used an LMS (Learning Management System) as a medium for course delivery for both offline and online offerings. Along with various student information management systems (SISs), LMSs are providing the educational

community with a huge amount of unexploited data about students' Learning performances, individualities, and behaviors.

2.1 Using clustering in EDM:

Researchers have [5] applied statistical clustering method like K-means clustering and Hierarchical clustering to student annotations. And they proved that by using these clustering methods, the creation of students with similar learning style cluster is improved and is faster. Association rules, clustering, classification, sequential pattern analysis, dependency modeling, and prediction have been used to improve web based learning environments to subsequently enhance the degree to which the educator can evaluate the learning process [4].The below table shows clustering algorithms used in EDM[3].

Table 1.Clustering algorithms used in EDM

Problem/ Objective	Algorithm/Method	Dataset/Data source
Evaluating undergraduate student academic performance. [7]	Using a combination of DM methods like (Artificial Neural Network), Farthest First method On the basis of k-means clustering and Decision Tree as a classification approach	Student data of the Computer Science at Faculty of Science and Defense technology, National Defense university of Malaysia (NUDM)
To analyze the web log data files of a Leaning Management System(LMS) [8]	Markov Clustering (MCL) algorithm for clustering the students activity and a Simple KMeans algorithm for clustering the courses	The dataset was collected from the Technological Education institute of Kavala uses the Open e-Class e-learning platform. The data are from the Department of Information management and comprise 1199 Students and 39 courses. The data are in ASCII obtained from the Apache server log file.
Deals with clustering of student access patt-erns in an e-learning environment [9]	Fuzzy sets and Transitive closure	No real life data used. Paper is based on hypothetical data

2.2 Moodle:

In 2013[11] Angela Bovo, St´ephaneSanchezy, Olivier H´eguyz and Yves DuthenX developed a system which predicts students' performance during an online curriculum on a LMS Experiments in clustering were carried out using real data obtained from various courses dispensed by a partner institute using a Moodle platform. They have compared several classic clustering algorithms on several groups of students. A professional training institute named Juriscampus provided real time data. The system collects and analyses data from any LMS with a logging system. Moodle's logging system keeps track of what materials students have accessed and when and stores this data in its relational database.

The **limitation** of this system is that we could not see qualitative difference among the students' activity but quantitative. And when the numbers of clusters were too large, clusters contain only one student.

2.3 SARG [10]:

It is an open source tool that allows you to analyze the squid log files and generates useful reports in **HTML** format with information about users, IP addresses, top accessed sites, total bandwidth usage, elapsed time, downloads, access denied websites, daily reports, weekly reports and monthly report . The **limitation** of system is that it only generates Static report.

III. MODELING AND METODOLOGY

S= {I, O, P}

Where, I= Input, O=Output, P=Processing

Modules in the system:

S= {Report Generation, Domain wise grouping}

A. Report Generation:

I= {dt, url, tim, dept, stud}

Where,

dt= Date, url= Url name, dept= Department, stud= Student Id or Name

P= {$\sigma_{date=dt}$ (**tblLog**),$\sigma_{URL=url}$ (**tblLog**),$\sigma_{time=tim}$ (**tblLog**), $\sigma_{department=dept}$(**tblLog**), $\sigma_{name=stud}$ (**tblLog**)}

O = {Report Generated}

B. Domain wise Grouping:

I= {Domain name},*P*= {*K*-Means Clustering}, *O*= {Grouping according to domain}

IV. PROPOSED SYSTEM

The thought behind the idea is to improve qualitative analysis and to boost student's interest and show them right way. As described earlier, the paper is about developing system in the form of a web application. It is a kind of analyzer where tutors can quickly analyze the interest of their student.

The main Objective is that:

- To study students' behavior and provide them a proper guidance related to their academics.
- To predict future results of students' and ultimately the success of an educational organization.

The figure given below shows that the proposed system architecture. Student's access internet and that web access log is maintained in a database. And simultaneously their academic performance (percentage marks) is maintained in database. Then clusters of students are formed who shows similar behavior. Finally, teachers access those clusters and find students according to the date, time and day of accessing internet, also tutor can find students who have accessed particular websites, also one can group them based on downloading size.

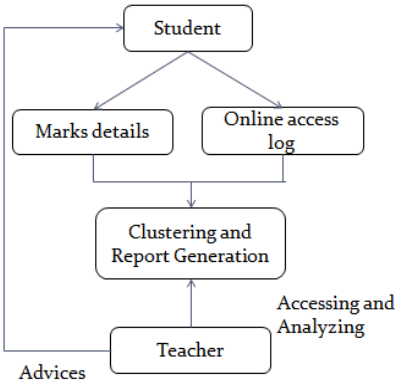

Figure 1. Proposed system architecture

V. DATA SET AND ALGORITHM

A. Data set:

The sample collected data is student1s' real time web accessed data and their details of different subject marks in semester wise have been recorded and applied the data mining process on that.

Table 1. Data set

Attributes	Data type	Description
Day	Varchar	Day of accessing
Date	Integer	Date of accessing
Month	Integer	Month of accessing
IP address	Varchar	IP address of the system
URL	Varchar	url accessed
User name	Varchar	Username of the user
Downloading	Integer	Data Usage

B. Methodology

Clustering: Given a data base D= {t1, t2….tn} of tuples and an integer value k, the clustering problem is to map f: D→{1…., k} where each ti is assigned to one cluster Kj, $1 \leq j \leq k$. A cluster Kj contains precisely those tuples mapped to it that is Kj= {ti / f (ti) =k, $1 \leq i \leq n$, and ti in D}

***Algorithm 1: K*-Means Algorithm**

N: Number of Data Object, K: Number of Clusters, Obje [N]: array of data objects
Clusters [K]: array of Cluster centers ,member [N]: array of Object membership
kmeans_clustering ()

```
1 While ð/N >threshold,
2 ð← 0
3 for i← 0 to N-1
4 for j← 0 to K-1
5 dist← | Obje[i] – Clustering[j] |
6   if dist<min_dist
7  min_dist←dist
```


8 n ← j
9 if member[i] ≠ n
10 ∂←∂+1
11 member[i] ← n
12 new_cluster[n]←new_cluster[n]+obje[i]
13 new_cluster_size[n]←new_cluster_size[n]+1
14 for j← 0 toK-1
15 cluster[j][*]← new_cluster[j][*]/new_cluster_size[j]
16 new_cluster[j][*] ← 0
17 new_cluster_size[j] ← 0 [6].

CONCLUSION

This paper is basically concentrating on educational data mining. It shows the data mining can be used in educational fields particularly to improve students' performance. That is result in forming the group of students who shows strong correlation between mental condition and their final academic performance. This will help to improve the performance of students. Ultimately, this will be useful for teachers to monitor students.

REFERENCES

[1] Han Jiawei, Micheline Kamber, "Data Mining: Concepts and Technique" Morgan Kaufmann Publishers, 2000.
[2] W.M.R. Tissera, R.I. Athauda, H. C. Fernando "Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining", IEEE International Conference on Information Acquisition,2006
[3] Ashish Dutt, Saeed Aghabozrgi, Maizatul Akmal Binti Ismail and Hamidreza Mahroeian "Clustering Algorithms Applied in Educational DataMining"*2, March 2015*
[4] B.Liu and J.-H.Zhao, "Non-linear correlation techniques in educational data mining," in *Proc. 2009 Sixth Int. Conf. Fuzzy Syst.Knowl.Discov.*, 2009, pp. 270-274.
[5] M. Wook, Y. H. Yahaya, N. Wahab, M. R. M. Isa, N. F. Awang, and H. Y. Seong, "Predicting NDUM student's academic performance using data mining techniques," in *Proc. 2009 Second Int. Conf.Comput. Electr. Eng.*, 2009, pp. 357-361.
[6] Md. Hedayetul Islam Shovon*, Mahfuza Haque "Prediction of Student Academic Performance by an Application of K-Means Clustering Algorithm" 7, July 2012
[7] T. Peckham and G. McCalla, "Mining student behavior patterns in reading comprehension tasks," *Int. Educ. Data Min. Soc.*, pp. 87-94, 2012.
[8] S. Chen and X. Liu, "An integrated approach for modeling learning patterns of students in web-based instruction: A cognitive style perspective," *ACM Trans. Comput. Interact.* vol. 15, no. 1, 2008.
[9] C. Romero, S. Ventura, and E. Garcia, "Data mining in course management systems: Moodle case study and tutorial," *Comput.Educ.*, vol. 51, no. 1, pp. 368-384, Aug. 2008
[10] www.techmint.com
[11] Angela Bovo, St´ephane Sanchezy, Olivier H´eguyz and Yves Duthen "Clustering Moodle data as a tool for profilingstudents"2013 IEEE.

