Extending the Student's Performance via K-Means and Blended Learning

Zahida Parveen
Department of Computer Science
University of Hail, Hail
Saudi Arabia

Anushya Alphones
Department of Computer Science
University of Hail, Hail
Saudi Arabia

Samina Naz
Department of Computer Science
University of Hail, Hail
Saudi Arabia

Abstract— **In this paper, we use the clustering technique to monitor the status of students' scholastic recital. This paper spotlights on upliftment the education system via K-means clustering. Clustering is the process of grouping the similar objects. Commonly in the academic, the performances of the students are grouped by their Graded Point (GP). We adopted K-means algorithm and implemented it on students' mark data. This system is a promising index to screen the development of students and categorize the students by their academic performance. From the categories, we train the students based on their GP. It was implemented in MATLAB and obtained the clusters of students exactly.**

Keywords- student's performance; k-means clustering; Blended Learning; gifted students.

I. INTRODUCTION

Now a day we are in age that is referred to an information age. In this age, we admit that information leads to power and success. This information age there is a vast amount of data available. It is necessary to extract useful information to analyze these large volumes of data. Determine hidden patterns from huge sets of data sources, is known as knowledge discovery and data mining (KDD). Data mining is quicker and easier, with good tools. Data mining was urbanized demerit depending on a statistician and to assist business people to form useful discoveries from data independently. Data mining (KDD) process are to understand the application domain and Identify data sources and select target data.

Data Mining (DM) applying in education is an emerging interdisciplinary research field referred to as "EDM" educational data mining, anxious with developing, researching, and implement computerized methods to detect the unique types of data from enormous volume that discover from educational environments. Its aim is to well recognize how students learn and understand the framework in which they learn to gain acumens for increase educational results. For solving the scholastic environments massive amount of potential data disputes, the traditional DM techniques cannot be applied directly to these types of data and problems. Some specific DM methods and the knowledge discovery process have to be adapted.

Clustering is applied to position elements of a database into specific groups according to some attributes. Data Clustering is technique to analyze the unsupervised statistical data. It is used to categorize the similar data into a uniform group. It is used to work on a large data-set to determine hidden pattern and association that helps to make decision rapidly and proficiently [1]. In a statement, Cluster analysis is used to segment a big set of data into subsets named clusters. Each cluster is a group of data objects that are related to one another are positioned within the same cluster but are different to objects in other clusters. In order to exercise clustering, we referred to one of the most frequently used algorithm that is the K-Means.

With the huge burst of internet the infusion of web-centered technologies into the learning and teaching process is obvious and resulted in the trend towards E-learning. Functionality of E-learning can be used as sustenance in the process of learning- teaching as it is a substantial tool for fast conveyance, improved communication, and assistance of both instruction and learning methods[2]. The useful information generated by the educational data mining can be better utilized when the learning process becomes computerised, to enhance the learning model for academic purposes. The incorporation of educational Data Mining techniques into E-learning environment have been produced successful results as demonstrated in several studies. The application of data mining techniques and concepts in E-learning systems helps to support educators to improve the E-learning environment. Most of the higher level education sectors have adopted online learning to support the students, and fulfill their demands for a flexible and expedient way to support them in their learning keeping the quality at same level.

University of Hail the chosen entity for this study is currently using Blackboard, as their Learning Management System (LMS).A Learning Management System (LMS) is an online portal that connects lecturers and students and to promote group-based collective learning and teaching.

In this paper an unsupervised data mining technique, clustering has been used for finding clusters of students with similar level of grades achievement. The successful analysis of mining results will enable teachers, students and the course content designers to improve in their respective circle of action and to better allocate resources and organize the learning process in order to improve the learning experience.

Rest of the paper is organized as follows next section II discusses the review of related literature. Section III elaborates about the K-Means algorithm and clustering the students, Section IV presents data source, Section V describes experimental analysis performed with dataset and proposal is presented in next. Finally, Section VII concludes the paper and hopeful tracks of future enhancements.

II. LITERATURE REVIEW

There are many data mining techniques widely adopted by many authors in their research blending with e-learning. In [3], studies on how Data mining methods can advance learning potentials in e-learning atmosphere effectively. The technique used in this paper based on data clustering which is recommended to be enhancing collaborative learning and to contribute incremental student judgment. The same clustering technique has been utilizes to group similar course material which can assist the e-learners to easily find and classify distributed courseware assets. To cluster related learning materials, an aspect of clustering tool used in the implementation of the Bisection K-Means algorithm.

In [4], data clustering technique was proposed for student judgment and stimulate group-based cooperative learning. The research offer many possible solutions for the problems faced in distance learning by the use of Web Mining techniques. Many areas of data mining provide aids to improve e-learning as an excellence education technique. Using Classification Models of data mining student progress, teachers" performance and student behavior can be predicted. By using clustering method a model can be built to improve learning process.

The system [5] used Weka system that has many clustering algorithms. The specific algorithm used in [5] is one of the simplest and most popular algorithms that are K-Means. K-Means clustering has been used to group the students who are taking some specific courses by building the clusters depending on the activities that they have performed on Moodle, during the whole course length and the final grades. The K-Means algorithm executed with a value of 3 to the number of clusters. The said system provides information about the cluster centroids of each cluster, the number and percentage of instances in each cluster[6]. The main aim behind this study was to divide the students into similar capabilities groups and hence apply on them appropriate teaching methods and techniques.

The study of system [7] revealed that the authors utilized data mining technique to discover knowledge. All the available data was collected that is also include the frequency of Moodle e-learning facility. Mainly they discovered implication rules and used lift metric to sort the rules afterwards these rules were

visualized by the concerned teams. The rules of classification were discovered using decision tree. The writers adopt EM-clustering to group the students. Outlier analysis has been utilized to identify all outliers in the data. To improve the performance of students, all of this knowledge can be used.

In [8] the very obvious property of clustering, which is a core task of data mining has been took in to consideration that is finding the groups of objects. These objects from the same cluster have similar properties as compare to the objects from other clusters.

III. K-MEANS CLUSTERING

K-means algorithm is functional for undirected information findings and is relatively simple. K-means has set up wide spread usage in lot of fields. It accepts the number of clusters, k to group data into, and the dataset to cluster as input values. Then it creates the first K initial clusters (K= number of clusters needed) from the dataset by choosing K rows of data randomly from the dataset. K is positive integer number. The grouping is done by minimizing the sum of distances between data and the corresponding cluster centroid. Thus, the purpose of K-mean clustering is to classify the data.

A. K-Means Clustering Algorithm

In this section, the steps of K-Means algorithm are explicated here.

The K means algorithm will do the three steps below until convergence

Iterate until stable (= no object move group):

1. Determine the centroid coordinate

2. Determine the distance of each object to the centroids

3. Group the object based on minimum distance (find the closest centroid)

K-Means works very well, if variables are huge, then K-Means computationally faster than hierarchical clustering, if K is small and K-Means produce tighter clusters than hierarchical clustering, especially if the clusters are globular[9]. But, on the other hand, difficult to predict K-Value, with global cluster, it didn't work well, Different initial partitions can result in different final clusters, and it does not work well with clusters (in the original data) of Different size and Different density.

For analyze the student performance, we use the GPA as measurement in 9 different aspects which is explicitly shown in Table 1. In this work, the value of K is sticking to 9 and we receive the maximum of 9 clusters as the output also the centroid is predefined as shown in Table 1. From the 5 clusters, faculties can identify the slow learners, poor concentrated students and train them in a proper way by the remedial plans and actions.

B. Measurement of final grade

The students" final grades, submitted to the University Registration Department by each course instructor, were

supplied to us by the Registration Department. In the actual data analysis, the final grades were categorized into nine groups as shown in Table-1.

TABLE I.   CONVERSION MARKS TO GRADE

GPA	Grade Code	Grade	Percentage	Centroid
4.00	A+	EXCEPTIONAL	95-100	97.5
3.75	A	EXCELLENT	90-94	92
3.50	B+	SUPERIOR	85-89	87
3.00	B	VERY GOOD	80-84	82
2.50	C+	ABOVE AVERAGE	75-79	77
2.00	C	GOOD	70-74	72
1.50	D+	HIGH PASS	65-69	67
1.00	D	PASS	60-64	62
0.00	F	FAIL	< 60	29.5

IV. DATA SOURCE

The data source used for this study is dataset of students in Department of Information and Computer Science, College of Computer Science and Software Engineering has been considered for clustering. A total of 20 records with 4 attributes were used for clustering. Figure 1 lists the attributes.
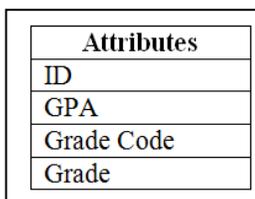

Figure 1.   Attributes of student dataset

V. EXPERIMENTAL ANALYSIS

Experiments are conducted with MATLAB. Data sets of 20 records with 4 attributes are used. To enhance the prediction of students „performance, K- means clustering is incorporated with faculties. Experimental analysis of K-means is exhibited in Figure 2 and Figure 3. Observations show evidences of that the K-means clustering technique outperforms for clustering.
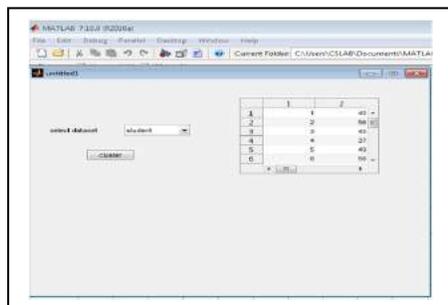

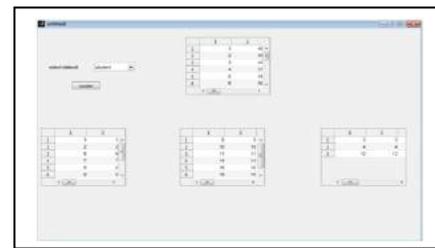Figure 2.   Screen shot for loading the student's dataset

Figure 3.   Screen shot for clustering the student's dataset

Graded Point Average (GPA) is a universally applied marker of academic performance. So, GPA still remains the most common factor used by the academic planners to evaluate progression in an academic environment. The main purpose of clustering is to divide the students into similar kind of groups according to their distinctiveness and capabilities (Kifaya, 2009). This work can facilitate both instructor and student to improve the education excellence. This study makes use of cluster analysis to divide students into groups according to their GPA. It is used to classify the same data into a homogeneous group. Cluster analysis is used to segment a large set of data into subsets called clusters[10]. Each cluster is a collection of data objects that are similar to one another are placed within the same cluster but are dissimilar to objects in other clusters. Output yields 9 clusters of students who are Exceptional, Excellent, Superior, Very Good, Above Average, Good, High Pass, Pass and Fail. In this study, we termed the students who are in exceptional group and Fail as Gifted students and Dunce students interchangeably.

VI. PROPOSAL

Through e-learning, these two group will be monitored via Blackboard and to sound more advanced, encourage them to present research papers, make a conscious effort to enhance their knowledge for gifted students and give remedy for the dunced students through Differentiated Instruction, Scaffolding, Graphic Organization, Mnemonics, Multisensory Instruction. It can be applied for CCSE (College of Computer Science and Software Engineering) in university of Hail in future. The system flow architecture is given in Fig-4.
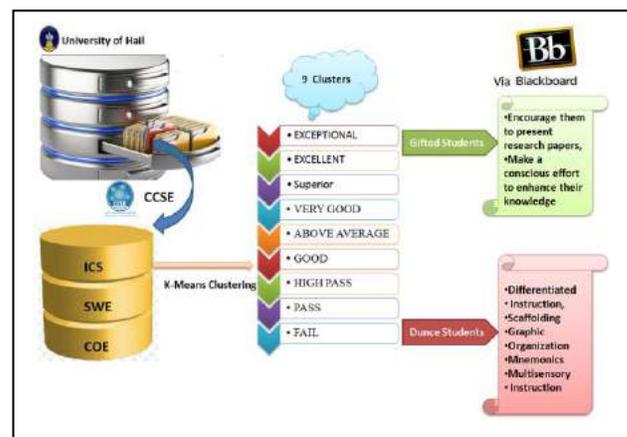

Figure 4.   System Flow Architecture

VII. CONCLUSION

We attach this idea to obtain higher accuracy to predict the students" performance by K-means algorithm which provides the best result in 3 (maximum of 9) clusters, because there is no student belongs the remaining clusters. It would be a promising technique for predicting the students in present scenario. This exploration gives a hand in the complexity of identify the students.

In the future, this work can be expanded by to predict the students who are almost likely to drop out the course early and also to group the students into specific categories that can be targeted with personalized interventions if it is predicted that drop out is imminent.

REFERENCES

[1]   "On-Line Clustering Of Lecturers Performance Of Computer Science Department Of Semarang State University Using K-Meansalgorithm", ENDANG SUGIHARTI1, MUCH. AZIZ MUSLIM2, Journal of Theoretical and Applied Information Technology10th January 2016. Vol.83. No.1

[2]   "An Approach of Improving Student's Academic Performance by using K-means clustering algorithm and Decision tree" , Md. Hedayetul Islam Shovon, Mahfuza Haque, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.3, No. 8, 2012

[3]   Margo, H.: Data Mining in the e-Learning Domain. Computers & Education (2004)

[4]   Tang, T.Y., McCalla,Smart Recommendation for an Evolving e-Learning System:Architecture and Experiment. International Journal on e-Learning (2005)

[5]   Romero, C. et al., Data mining in course management systems: Moodle case ..., Computers & Education (2007), doi:10.1016/j.compedu.2007.05.016

[6]   "Clustering In RapidMiner", by Anthony Moses Jr on 23 April 2014

[7]   Mining Students Data to Analyze Learning Behavior: A Case Study, ALAA EL-HALEES Department of Computer Science, Islamic University of Gaza P.O.Box 108 Gaza, Palestine alhalees@iugaza.edu.ps

[8]   Mining Educational Data to Improve Students' Performance: A Case Study Mohammed M. Abu Tair, Alaa M. El-Halees, Volume 2 No. 2, February 2012 ISSN 2223-4985 International Journal of Information and Communication Technology Research, ©2012 ICT Journa

[9]   "Face Recognition System for Set-Top Box-Based Intelligent TV", Won Oh Lee, Yeong Gon Kim, Hyung Gil Hong and Kang Ryoung Park , Sensors 2014, 14, 21726-21749; doi:10.3390/s141121726

[10] "RESUMEN DE TESIS DOCTORAL Restricted Conceptual Clustering Algorithms based on SeedsAlgoritmos Conceptuales Restringidos basados en Semillas", Computación y Sistemas Vol. 11 No. 2, 2007, pp 174-187 ISSN 1405-5546

AUTHORS' PROFILE

Zahida Parveen has completed MCS (Master in Computer Science) NUML Islamabad, Islamabad, Pakistan. B.Sc (Computer) Punjab University, Punjab, Pakistan. She is currently working as lecturer of computer science in University of Ha'il from year 2010. Her research interests are Data Mining,Clustering, E-Learning, Cloud Computing, Programming Languages, Software Engineering and Big Data.

Dr A. Anushya received the Bachelor of Computer Science (B.Sc.,) degree from the Mother Theresa University, Kodaikanal, India, in 2006 and the Master of Computer Applications (M.C.A.) degree from the Bharathidasan University, Tiruchirapally, India, in 2009. She received the Ph.D. degree from Department of Computer Applications, Manonmaniam Sundaranar University in 2016, Tirunelveli, India. Her research interests include Data Mining, fuzzy, Rough set, Soft computing, Neural Network, Image classification and clustering and Artificial Intelligence

Samina Naz has completed MCS (Master in Computer Science) NED University Karachi,Karachi, Pakistan. B.Sc (Computer) Islamia University Bahawal pur, Punjab, Pakistan. She is currently working as lecturer of computer science in University of Ha'il from year 2010. Her research interests are Data Mining,Clustering, E-Learning,Cloud Computing, Programming Languages, Software Engineering and Big Data.