Review Article

Tendency of Educational Data Mining in Digital Learning Platform

Gajendra Sharma*

Department of Computer Science & Engineering, Kathmandu University (KU), Nepal

***Corresponding author:** Gajendra Sharma, Department of Computer Science & Engineering, Kathmandu University (KU), Nepal*

Abstract

With the advancement of technology learning process have been more reachable and interactive like never before. Till the date many online learning platforms have been introduced. The platforms with constant improvisation in teaching-learning technique have been able to sustain. For improvisation there need constant analysis of the data and implement the suggested changes are required. The analysis of the educational data is using data mining techniques is called Educational data mining. EDM helps to discover the patterns of learning behavior hidden in the data sets. This research paper aimed to do a review of different other papers which were based on EDM in online learning environment. It has been seen that most paper used classification and clustering as their data mining technique. K-means clustering has been used as cluster analysis technique for exploring the dataset. Similarly, Weka tools had been found to be used as data mining software. Future recommendations in EDM are presented in terms of future scope of the researches related to it to make the researches more trustworthy.

Keywords: Data mining; Educational data mining; Online learning; E-learning; Classification; Clustering; K-Means clustering

Introduction

Educational data mining (EDM) is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context [1]. Data mining is a process of identifying and extracting hidden patterns and information from huge repositories of data which are better known as databases and data warehouses [2]. During last decades the use of data mining have been widely used in the different fields like business, bioinformatics, marketing campaigns, cyber security, education, research and many more. Education is one of the fields which are most affected and facilitated by the advancement of the technology. Technology is major tool for learning nowadays which help students in active and interactive learning. The term active and interactive learning means online learning here. The analysis of data from online learning platforms discovers the effectiveness of the online learning method. The knowledge discovered from data mining can be used as a helpful and constructive recommendations in online learning to enhance their decision making process, to improve students' academic performance and to better understand students' behavior. The focus of Educational Data Mining in this research is learning resources and student behavior and performance. The research takes a look at different Educational Data Mining researches and the techniques they have used to effectively classify the educational as mentioned earlier. The aim of this research is to find out the kind of data researchers work on, tools and techniques used in data mining as well as the success and challenges of the EDM.

Literature Review

Data mining tools can include statistical models, mathematical algorithms and machine learning methods. The data mining techniques are able to find out information within the data that queries and reports can't effectively reveal [3]. Educational Data Mining is capable of revealing system usage behaviors using data mining techniques. EDM uses many techniques such as Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbor, and many others [4]. One of the most useful data mining tasks in online learning is classification. To discover capability student groups with similar characteristics and reactions to a particular strategy, to detect students' misuse or game-playing, to group students who are

hint-driven or failure-driven and find common misconceptions that students possess, to identify learner's motivation to lower drop-out rates, to predict/classify students when using intelligent tutoring systems are some of the educational objectives to use classification [5]. Clustering is one of the data mining techniques. It can be used to discover the new categories, which share the similar interest. In clustering the similar instances are grouped together. Among the different available clustering methods, K-Means algorithm is generally used to divide data into natural groups based on their behavior for a larger dataset. In the K-Means clustering method, the number of clusters, denoted by K is needed to be predefined to apply the technique. The biggest problem in K-Means is the finalizing of the optimum number of clusters. WEKA and R were used for this study due its availability to download as open source software and compatibleness with CSV files [6].

The objective of clustering is to find high-quality clusters such that the Inter-cluster distances are maximized and the Intra-cluster distances are minimized. The clustering method applied in the paper "Mining Educational Data to Improve Students' Performance: A Case Study" is the k-means; the objective of this k-means test is to choose the best cluster center to be the centroid. The k-means algorithm requires the change of nominal attributes into numerical [7]. The conference paper "Examining students' online interaction in a live video streaming environment using data mining and text mining" which uses E-learning as the platform of research with the objective of mining the student online assessment data used classification, clustering, and association rule analysis as the data mining task [8]. Hung and Zhang [9] have done a study on undergraduate students using the decision tree technique to propose a predictive model of user performance and reveal the students online learning behavioral patterns. They have indicated that the majority of the students were passive learners and only tend to access e-materials, but did not seek any peer collaborations. However, the few active learners showed a high performance level. They have proposed a decision tree for predicting performance. According to Ratnapala et al. [6], the aim of their research was to use EDM techniques to conduct a qualitative analysis of students' interaction with E-learning system. The total of 412 students (enrolled in instructor-led non-graded and graded courses) was being researched on access behavior in an E-learning environment. K-means clustering method divided the student population into five access groups based on their course access behavior. This research concluded that the difference in the learning environments could change the online access behavior of a student group. Among these groups, the least access group (NG- 41% and G-42%) and the highest access group (NG-9% and G- 5%) could be identified very clearly due to their access variation from the rest of the groups.

Research Problem

There are various reasons to choose data mining. The major reason is that data mining involves the use of data analysis tools that

helps to discover previously unknown, patterns and relationships in large data sets. In this research, the focus is on Educational data. Both researchers have academic backgrounds, and they aimed to find the best way to show that Educational Data Mining is important and helpful for educational Institutions in decision making.

The researchers sought to answer the following questions:

1.   What kind of data should we be working on?

2.   How educational data are collected from students (whether from use of interactive learning environments, computer-supported collaborative learning, or administrative data from schools and universities)?

3.   How useful data mining is in the educational analysis?

4.   What data mining tools and techniques are best to be used?

Findings from Literature Review

With introduction to different online learning platforms it has been more accessible for students to learn at desired way. To improve the learning environment and to cope with the desired pace of learning of the students it is important to analyze the data collected from those E-learning environments. The data mining of these data help in finding the patterns of online learning behavior and promotes the decision making in introducing innovative and more interactive way of E-learning to students. It also helps to find the way to improve the way of assigning course and teaching way to the instructors. For the successful Educational data mining, we should select the mindful set of data; find out the best technique to work with the collected data. The correct selection of the data mining tools and techniques ensure the success in the educational data mining. The techniques discussed in the literature review are some of the common data mining techniques used in the field of EDM. There are many more such techniques for EDM. The works discussed above found to use predominantly one technique thus there is need of hybridized techniques (more than one technique) which compensate and complement each other. This has already been used in other data mining field and now is the time for EDM too to take the benefit of the hybridized technique. The result expected from the EDM are better understanding of the current cause and effects involved in the online learning system. The results of these kind of the research have always been questioned their trustworthiness. Hence the research should be conducted in trustworthy scenario so that the outcomes can be implemented into the system for further improvement

From the literature review above it can be seen that the classification, clustering and decision tree are popular data mining techniques. K-mean clustering is considered to be the popular cluster analysis technique for exploring the dataset. It has been seen that the use of Weka and SPSS Clementine 12.0 made Data Mining easier. On the other hand also good to use other tools such

as R, Tanagra, Rapidminer, and Oracle Data mining. Utilization of different tools for contrast and comparison may help also in finding the best tool for the data that is being mined. The researches above mentioned suggest that poor quality of e-learning materials, language and technology incompetence, technical problems, system problems, inadequate facilities, isolation and inadequate support are some of hindrances that prevent the students from being the active E- learner. Similarly, the self-motivation of the students plays vital role in the success of the e- learning.

Conclusion

Hence, we conclude that the educational data mining in the online learning environment have been following some similar manner only differing in the data set which raises the trustworthiness of the result of the studies. Therefore it's high time to discover different other technique in the EDM. One of the best ways is to use more than one technique together to contrast and compensate the pros and cons of the techniques used.

Future Scope and Recommendation

Here some of the suggestions to be considered for making results of the research more trustworthy:

a)   For the accurate outcomes the data sets are to be as large as possible. Similarly, the process of collecting data needs to be reorganized to show in data the true and sensible picture of the real world system.

b)   The researches show that only one technique was used by isolating from other techniques. There is a high demand to explore hybrid techniques or alternatives for conventional algorithms that can better perform the data mining.

c)   The data mining tool has to be integrated into the online learning environment as another author tool so that the data analysis goes on side by side in a single application. Feedback and results obtained with data mining can be directly applied to the e-learning environment.

References

1. Romero C, Ventura S (2010) Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 40(6): 601-618.

2. Romero C, Ventura S (2006) Data mining in e-learning. WIT press.

3. Silva C, Fonseca J (2017) Educational data mining: a literature review. In Europe and MENA Cooperation Advances in Information and Communication Technologies p. 87-94.

4. Brijesh Kumar B, Sourabh P (2011) Mining educational data to analyze students' performance. IJACSA 2(6): 63-69.

5. Romero C, Ventura S, Espejo PG, Hervás C (2008) Data mining algorithms to classify students. In Educational data mining.

6. Ratnapala IP, Ragel RG, Deegalla S (2014) Students behavioral analysis in an online learning environment using data mining. In 7th International Conference on Information and Automation for Sustainability p. 1-7.

7. Abu Tair MM, El-Halees AM (2012) Mining educational data to improve students' performance: a case study. Mining educational data to improve students' performance: a case study 2(2).

8. He W (2013) Examining students' online interaction in a live video streaming environment using data mining and text mining. Computers in Human Behavior 29(1): 90-102.

9. Hung JL, Zhang K (2008) Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. MERLOT Journal of Online Learning and Teaching 4(4): 426-437.

CTCSA
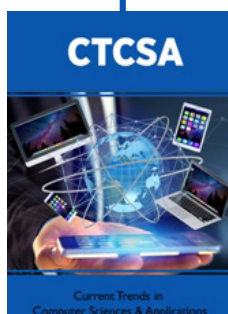
Current Trends in Computer Sciences & Applications

Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles