

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Education Data Mining: A Review

Joshita Goyal¹, Shilpa Sharma²

M.Tech Student, Department of Computer Science and Engineering, Lovely Professional University, Phagwara,
Punjab, India¹

Asst. Professor, Department of Computer Science and Engineering, Lovely Professional University, Phagwara,
Punjab, India²

ABSTRACT: Education data mining is emerging trend of research, so I have attempted to write a review paper describing its importance, the knowledge discovery process, popular algorithms used in this context and recent researches. I aim to enlighten the wide area of research that can be done in this area before writing this review paper. I have written the enhancements that can be done in the researches in context of EDM. This review paper will highlight the latest research that can be done in education data mining which will help improve our education system and will prove beneficial for the researchers who want to do research in this field.

KEYWORDS: Education data mining, knowledge discovery process, algorithms

I. INTRODUCTION

Education data mining is emerging area of research in technology which enables us to improve our education system to greater extent. Education data mining (EDM) comes under data mining which let us dig into the relation between various dimensions related to student performance. We can also call it as an approach by which we'll be able to provide better course material to students according to their interest which will help them excel in their specific field. In last decade a lot of researches has been done by researchers to bring up enhancements in the world of education and these are still on. Education data mining is integrated with big data and data analysis to unleash the relation between various attributes which are there in massive amount of data stored in databases of various institutions. Now time has come to utilize this data to unleash the hidden useful data which can be organized to form a useful knowledge[1].

Today education is not limited to classrooms but it has gone to some other level. Now there are various online courses available for the learners which can be joined by any student irrespective of age limit and at any time. Online videos and various tutorials available on YouTube and many learning platforms like MOOCs are available for learning. As education data mining has evolved from data mining so various techniques and tools used in data mining are all applied in EDM. In this review paper, we'll discuss the knowledge discovery process used in data mining, tools available in education data mining. In second section we'll have literature view illustrating certain researches which have been done in EDM. In third section we'll discuss that what enhancements can be done in existing researches thus giving an insight into future research followed by conclusion.

A. Education data mining

EDM can be called as a subset of a broad area of data mining. Its approach to make our education system a stronger and more productive. Prior to using education data mining we were having a lot of data which remain useless. But now with the evolution of education data mining we have been given opportunity to use that data to uncover the hidden information in it. With the techniques and tools we use in data mining we can bring up a robust system which is helpful

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

in making decision regarding student assessment and much more. As education is expanding its root from classrooms to electronic learning, more and more need of education data mining is required.

Education data mining is used not just in classroom and e-learning platforms but also in social sites. Education data mining has no limitations. It has its connection with each many users, students, teachers, institutions are some out of these. Thus it has connect whole world with it.

EDM in context of business: EDM works with data science, data analysis, big data and businesses. Electronic learning platforms are using EDM as a way to generate revenues in their business. They mine data and find ways they as how to improve their business and to attract more and more customers to it. This is key area where education data mining has shown its importance. By this education data mining is giving more and more employment opportunities to the individuals thus creating its importance. So here I have come up with latest researches in this context.

B. Knowledge discovery process

Education data mining follows the process of knowledge discovery which is a constitution of various steps. Here are different phases of KD process:

1. Cleaning: In this we clean data that is we remove unnecessary data from data and fill the missing data.
2. Integration: In this, we integrate data from various databases.
3. Selection: In this phase, we select data which is required for analysis.
4. Data transformation: In this data is transformed into form which will be applicable for mining process.
5. Mining operations: Now we'll apply various data mining techniques on it so that find hidden knowledge from it.
6. Result evaluation and presentation: Now result will be evaluated from previous operation and it will presented into readable form.

C. Popular algorithms used in Education data mining

There are various algorithms which are used in education data mining. Out of these several are under written.

1. Classification: It's algorithm which predicts that in group of items, a particular item will belong to which class. We can say that it will classify the items according to requirements. It uses classifiers.
2. Association: If in given data set we want to evaluate the relation between various dimensions then we can use this association algorithm.
3. Clustering: After classifying items, if you want to group these items in particular group then you can use clustering.
4. Regression: In order to find the relation between dependent and independent variables, regression is used. It gives numeric value of a dependent variable which is a function independent variables and set of parameters.
5. K-means: In case we are given a set of items and we want to create K groups of these and in each group set of items will have almost similar attributes.
6. Support vector machine: If you are given set of two types of items and the difference between two sets is clear enough to separate these items in two different groups so that the hyperplane is expressed as $y = mx + b$ then we use support vector machine
7. Apriori: It's a kind of association rule in which you can find out that which set of items occur more frequently.
8. K nearest neighbor : if you have labeled data then KNN will just store it. And if some new data is input the it will look for k neighbors to classify the data.

II. LITERATURE REVIEW

As education data mining has evolved into new area of researchers, therefore many researchers have worked to carve decision support system which can help education system in better dissemination of knowledge to the students

And objective is to have a better education system and better facilities of courses which can help students to achieve their goals. AnimeshGiri, M Vignesh V Bhagavath, BysaniPruthvi andNainiDubey have created a placement prediction system in which they predicted the possibility of undergraduate students who can get placed in a company using their academic information, programming skills, team work and analytical skills[2]. Researchers used k-nearest neighbor classifier for this and compared their results against other algorithms namely logistic regression and SVM and

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

found out which best suit their requirements[2]. Impact of poor teaching on students was evaluated by researchers Anupam khan and soumya K. Ghosh[3]. They used the concept of association mining in it. RohitAhlawat, SushilSahay, SaiSabitha and AbhayBansal worked to evaluate that what are the factors that affect the enrollment patterns in Indian universities using K-mean clustering[4]. In this they have considered various things like enrollment rates in universities, number of companies in each state, foreign students enrollment in various states and crime rates[4]. SubhashiniBhaskaranSailesh and Dr. Kevin j Lu collaborated to analyzes that whether a student can complete degree from higher institute within optimum time or not and for student who can complete their degree in optimum time, their decision support system will help them increase their scores[5]. Likewise Haiyang Xia, Jiaxin Han, Jie Kong, Wejuan Wei and Lei Zhang used contrast target rule extraction to help teachers know that which student seldom fail or always fail in exams and what inputs are required for them[6]. Other researchers like Larian M Nkomo and Muesser Nat invented a decision support system by taking data from e-learning platform Moodle that which student access resources at last minute and who usually procrastinate online submissions[7]. On the other hand El Harrack Othman, SlimaniAbdelali and El BouhdidiJaber presented better approach to mine web videos using metadata based clustering with better search results over traditional signal processing techniques[8].

Table 1. Researches in EDM

References	Authors	Objectives	Algorithms used
[2]	AnimeshGiri, M Vignesh V Bhagavath, BysaniPruthvi and NainiDubey	To create a placement prediction system to predict the possibility of undergraduate students who can get placed in a company	k-nearest neighbor classifier
[3]	Anupam khan and soumya K. Ghosh	To evaluate the Impact of poor teaching on students	association mining
[4]	RohitAhlawat, SushilSahay, SaiSabitha and AbhayBansal	To evaluate the factors affecting enrollment patterns in Indian universities	K-mean clustering
[5]	SubhashiniBhaskaranSailesh, Dr. Kevin j Lu and Mansoor Al Aali	To classify students according their performance so that they can be assisted with better advise to pass their degree within optimum time and with flying colors.	Genetic algorithm
[6]	Haiyang Xia, Jiaxin Han, Jie Kong, Wejuan Wei and Lei Zhang	To create a decision support system which can help teachers evaluate that which student seldom or always fail in exams	contrast target rule extraction
[7]	Larian M Nkomo and Muesser Nat	To evaluate students last minute access before exams to online resources and procrastination in assignment submission.	Apriori algorithm
[8]	El Harrack Othman, SlimaniAbdelali and El BouhdidiJaber	To provide a better approach to mine web videos over traditional approaches	Clustering and classification
[9]	Cana Tastimur, Mehmet Karakose and Erhan Akin	To simplify the application of accreditation system	K nearest neighbor

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

III. ENHANCEMENTS THAT CAN BE DONE IN EXISTING RESEARCHES

As we have discussed that education data mining researches is hot topic now a days and many researches have been in this context as mentioned above. But these researches have focused on limited data and limited attributes have been taken into consideration while creating some decision support system. In the time when we have the power of big data available with us then we need to utilize it so that that a more productive and fruitful system comes up which can assist us in improving our current education system.

A decision support system can be created which can provide us with the information what kind of knowledge should be disseminated to the students by not just considering their interest but also according to the company demands. By this skills of students can be improved to much greater extent and better quality professionals can be produced by the power of education data mining.

IV. CONCLUSION

This paper was an attempt to give an insight into education data mining and kind of researches which are going on in it's context. I've discussed various education data mining, knowledge discovery process involved in it, various data mining techniques that are frequently used in this, recent researches and enhancements that can be done in it.

From this I have concluded that it's difficult to estimate that how wide this topic is. As a future scope I will focus on to create a decision support system according to the difficulty level of subjects and placement requirements a decision support system which will assists student to schedule their tasks accordingly so that there is minimum burden on them of studies and they work efficiently in their daily routine. This will help them achieve their goals.

ACKNOWLEDGMENT

I would like to thank Lovely professional University for extending their help in writing this review paper and to guide me to write on education data mining.

REFERENCES

- [1] Sagardeep Roy, Dr. Shailendra Narayan Singh, "Emerging Trends in Applications of Big Data in Educational Data Mining and Learning Analytics," *2017 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence*, (2017)193-198, 10.1109/CONFLUENCE.2017.7943148
- [2] AnimeshGiri, M Vignesh V Bhagavath, BysaniPruthvi and NainiDubey, "A placement prediction system using K-nearest neighbors classifier" *2016 second international conference on cognitive computing and information processing (CCIP)*, 2016, 1-4, 10.1109/CCIP.2016.7802883.
- [3] Anupam khan and soumya K. Ghosh, "Analysing the impact of poor teaching on student performance" *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 2016, 169-175, 10.1109/TALE.2016.7851789
- [4] RohitAhlawat, SushilSahay, SaiSabitha and AbhayBansal, "Analysis of factors affecting enrollment pattern in indian universities using k-means clustering", *2016 International Conference on Information Technology (InCITE)-The next Generation IT Summit on the theme-Internet of Things: Connect your Worlds*, 2016, 321-326, 10.1109/INCITE.2016.785763.
- [5] SubhashiniBhaskaranSailesh, Dr. Kevin j Lu and Mansoor Al Aali, "Context driven data mining to classify students of higher educational institutions", *2016 International Conference on Inventive Computation Technologies (ICICT)*, 2016, 1-7, 10.1109/INVENTIVE.2016.7824869
- [6] Haiyang Xia, Jiabin Han, Jie Kong, Wejuan Wei and Lei Zhang, "Discovering the academic situation of students by relationship mining", 2016, 189-195, 10.1109/ES.2016.31
- [7] Larian M Nkomo and MuesserNat, "Discovering students use of learning resources with educational data mining," *2016 HONET-ICT*, 98-102, 10.1109/HONET.2016.7753428.
- [8] El Harrack Othman, SlimaniAbdelali and El BouhdidiJaber, "Education data mining: Mining MOOCs Videos using metadata based approach," *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*, 2016, 531-534, 10.1109/CIST.2016.7805106
- [9] Cana Tastimur, Mehmet Karakose and Erhan Akin, "Improvement of relative accreditation methods based on data mining and artificial intelligence for higher education," *2016 15th International Conference on Information Technology based Higher Education and Training (ITHET)*, 2016, 1-7, 10.1109/ITHET.2016.7760736