Applying Data Mining for Framing of Computer Science Curriculum

Kavita OZA
Shivaji University, Kolhapur, India
skavita.oza@gmail.com

Rajanish KAMAT
Shivaji University, Kolhapur, India
raj_kamat@yahoo.com

ABSTRACT

The main objective of higher education is to enhance learning capabilities of students. In any standard scenario of tertiary education, students have to learn and undergo many courses to get their postgraduate degree. Therefore, while designing the curricula care has to be taken to offer appropriate courses in a semester so as to enhance the grasping ability of the student. This above mentioned situation is more fitting to disciplines like Computer Science where a Mastersprogramme may have more courses in a semester with overlapping or complimentary contents. In the above context the present paper has investigated the arrangement and rearrangement of the subjects in two postgraduate academic programmesvz. Masters in Computer Science and Masters in Computer Applications being conducted at Department of Computer Science of Shivaji University, Kolhapur. Data Mining specifically, data clustering techniques were used for mining the relationship between the academic programmes wherein the subjectmarks(result) were taken as the key metrics. The paper details out the findings and focuses on efficient combination of courses to be offered at each semester to improve students' learning behavior which can also serve as a sort of guidelines to curriculum framing bodies of the University.

Keywords: *Education data mining, coursedesign, K-means, clustering, association mining*

INTRODUCTION

There are increasing research interests in using data mining in education. Educational data mining deals with development of algorithms to discover knowledge from voluminous data originating from various instructive environments. Due to this voluminous data there is need to extract some useful information to improve thequality of managerial decisions to deliver quality

education. Looking at the rate of obsolescence of knowledge in the areas of Computer Science, framing the curriculum is really changeling. Curriculum framing bodies often finds it difficult to cram the voluminous course content in the limited contact hours. With the emerging model of 'Choice based credit System' (CBCS) the syllabus to some extent can be enriched by giving freedom to the learners to select the course themselves based on their pace of learning. Another global trend to enrich the course structure by using the 'Massive Open on-line Courseware' (MOOC) has also come to the rescue of the curriculum framers for uncompressing the contents of the syllabi by offering few course in on-line mode. However, the models such as CBCS or MOOC adopted from west have not matured enough in the context of the developing countries like India. Therefore still there is lot of gaps in the research on curriculum framing by adopting the data mining approach. Through the present paper the authors highlight the application of data clustering technique which is a subclass of data mining to help the curriculum framers to decide about the fate of the courseware with overlapping content. The paper also analyses and put forths the issues as regards to whether the course wares can be treated as prerequisites so as to enhance the learners grasping abilities.

PRIOR ART

Data classification can also be used to classify students based on data collected from their social behavior to predict probable drop-outs in academics (Bayer, Bydzovská, Géryk, Ob\vsıvac, & Popelınskỳ, 2012). To improve graduate students' performance, and overcome the problem of low grades of graduate students, many data mining techniques are applied viz. association, classification, clustering and outlier detection rules can applied at various stages of mining (Tair & El-Halees, 2012). Data classification using decision tree is one of the technique which helps in early identification of dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling (Baradwaj & Pal, 2012). A case study of a university has been carried out to improve the quality of education by analyzing thedata and discover the factors that affect the academic resultsso as to increase success chances of students (Kumar & Chadha, 2012). Another classification technique called Bayesian Classification is also used to predict the students who are doing well in the academics (Pandey & Pal, 2011). A focus on developing new tools and application of these tolls for discovering patterns in data to predict next appropriate action to be taken is given by (Bienkowski, Feng, & Means, 2012).

A lots of work has been carried in mining useful patterns from educational data. Proposed work uses clustering as one of the data mining tool to predict the learning behavior of students. A simple K-means clustering algorithm has been used. It unsupervised learning algorithms. The algorithm classifies a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids are chosen randomly. he next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is

completed. Again data from newly formed cluster is used to calculate new centroids again. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. An objective function like Euclidean distance is used to calculate the similarity distance.

APPLYING DATA MINING FOR COMPUTER SCIENCE COURSES

Department of computer Science of the ShivajiUniversity, Kolhapur (http://csd.unishivaji.ac.in/event/home.aspx) offers two post graduate course viz. Master of Computer Application (MCA) and Master of Science (Computer Science). Both courses have papers related to Data structures and Theory of computation. At M.Sc. course both Data structures and Theory of Computation papers are offered at the same semester and at MCA course they are offered at different semesters. Semester results of MCA and MSc. were analyzed with focus on Theory of computation paper. Two data sets were prepared viz. MCA_result and MSc _result having marks of MCA and M.Sc students of Theory of Computation. Clustering technique was used to analyze the data of both the data sets. A simple k-means algorithm was applied using WEKA (aopensourcedata mining tool) and following tables shows the result.

Table 1: Clustering data of MScRresult

Cluster no.	1	2	3
centroids	29	50.28	62.46
Size	16	32	13
Percentage	22%	55%	21%

Table 2:Clustering data of MCA_result

Cluster no.	1	2	3
Centroids	54.56	67.40	38.5
Size	25	22	14
Percentage	25%	22%	14%

Three clusters are generated for each data sets as detailed in the tables 1 and 2. The analysis pertaining to these data clusters has been covered in the following section.

RESULTS AND DISCUSSION

Table 1 reveals that there are 16 students who belong to cluster-1 having average marks of 29 out of 100. This cluster represents who have performed badly at their exams. Cluster-2 has 32 students with average marks of 50.28 out of 100. This indicate this cluster has average performers in the exam. And Cluster-3 has 13 students who are good in their performance at exams with average marks of 62.46. This class has failure rate of around 22%.

Analysis of the table 2, discovers that cluster-1 has 25 students with average marks of 54.56 who are average performers and cluster-2 has 22 students with 67.40 marks who are good performers at examinations. Cluster-3 has 14 students with average marks of 38.5 who are fair performers not bad if passing head is 35.

CONCLUSION

The results lead to the following conclusion:

1. MCA students study Theory of computation after completing their data structure course at previous semester. They have background of Data structure. Due to this background their performance at Theory of computation is better than MSc. Students.
2. Failure rate of MSc students is more than that of MCA students as M.Sc. students study the paper along with data structure in the same semester.
3. Number of students getting first class with distinction is more in MCA as compared to M.Sc. and performance of average performers is also better than M.sc i.e 54.56 as compared to 50.28.

Above results also indicate that students who have already studied data structure at previous semester do better in theory of computation than those who undertakes the same at the same semester.

The above findings further helps in better design of curricula by identifying the sequence of papers to be offered at semesters. This sequencing will help in enhancing the learning ability of students which ultimately will enhance the results. Study carried out was based on only two subjects but the authors are extending the research further to include more and thus form strategic direction for the entire curriculum of M.Sc. as well as M.C.A. Computer Science.

REFERENCES

Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.3417*. Retrieved from http://arxiv.org/abs/1201.3417

Bayer, J., Bydzovská, H., Géryk, J., Ob\vsıvac, T., & Popelınskỳ, L. (2012). Predicting drop-out from social behaviour of students. In *Proceedings of the 5th International Conference on Educational Data Mining-EDM 2012* (pp. 103–109). Retrieved from http://educationaldatamining.org/EDM2012/uploads/procs/Full_Papers/edm2012_full_12.pdf

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *Washington, DC: SRI International*.

Kumar, V., & Chadha, A. (2012). Mining Association Rules in Student's Assessment Data. *International Journal of Computer Science Issues*, *9*(5), 211–216.

Pandey, U. K., & Pal, S. (2011). Data Mining: A prediction of performer or underperformer using classification. *arXiv preprint arXiv:1104.4163*. Retrieved from http://arxiv.org/abs/1104.4163

Tair, M. M. A., & El-Halees, A. M. (2012). Mining Educational Data to Improve Students' Performance: A Case Study. *International Journal of Information*, *2*(2). Retrieved from http://esjournals.org/journaloftechnology/archive/vol2no2/vol2no2_7.pdf