A Data Mining Approach To Analize Academic Performance Across Various Dimensions Through Hierarchical Navigation

Dr. Preetham Kumar, Mithun Raman and Teslin Jacob
Department of Information and Communication Technology,
Manipal Institute of Technology, Manipal, India

Abstract—The economic success of every country depends on the quality of education imparted to its citizens. Hence education plays a vital role in building up any nation around the world. There is huge academic performance related data available, but they directly do not provide any useful information to educational managers and industry recruiters. It is often a challenging task to compare student performance and inter-disciplinary ranking. Through this paper, we demonstrate how data mining approaches can be used to compare and contrast student performance across various dimensions, and inter-department performance and anal-ysis. Collective assesment of individual student performance can decipher information which can be used at various levels and situations for better decision making support. Knowledge mined from the huge pool of academic performance related data can be used by industry professionals at the time of recruitment to select the best candidates based on their requrements. It can also be used by academic professionals to compare the performance across the various departments and across universities. Above all, a waste of effort, time, and money can be avoided accordingly for both students, recruiters and educators.

Index Terms—Data mining; data-driven decision support sys-tem; educational data mining; Student Profiling;

I. Introduction

Regarding human resource development, education always plays an important role in building up every country all over the world along the time. Educational decisions made by educational managers are important and have a strong impact on not only individual students and educators but also our society. If an educational decision were inappropriate, a lot of student and educators effort, time, and money would be wasted and bad outcomes would be produced and exist together with our lives for a long time [1].

The data in any given educational institution is growing rapidly. There is a need to transform this data into useful information and knowledge; hence we make use of data min-ing. Data mining is the process of extracting hidden, unknown and potentially useful information and patterns from databases, data warehouses or other such data repositories. Data mining methods can usually be divided into two categories: The first category is statistical, and its technology used probability analysis, relevance, cluster analysis and discriminated analysis; the other is machine learning in the artificial intelligence-based, through training and learning a large number of samples that need to set the mode or parameters. Educational data mining is the area of science where various methods are being developed for making discoveries within data. This data is obtained from an educational background. These methods provide an insight into a students behavioral patterns and the environment in which they learn [2].

In this paper, we show how knowledge derived from aca-demic data can be presented to academic professionals and industry recruiters for analysis and decision making support. Data can be mined to show the performance of students in subjects year-after-year. Data can also be abstracted to find the domain areas in which a particualr student has excelled; this data can be used by industry recruiters to shortlist students who are proefficient in a particular field of study. Academic professional can use the knowledge mined from the data to compare the performance of a particular department across various universities. Faculty evaluation can also be done using this data. Once information has been mined out of a large pool of data, it can be presented to the information-seeker in a way that will be useful to him. It is now just a matter of formatting the information to the users' preference. Once data has been extracted and formatted, we can choose what to present to the user based on his requirements. The data can be combined in different combinations which can then produce more valuable information without further mining data from the information.

II. Related Works

Data about studying and lecturing activities in educational organizations have been accumulated day by day over the time line. If this volume of educational data is explored and knowledge discovered from such data is utilized, studying and lecturing activities could be improved and be of much benefit to students, lecturers, educational organizations and society. So far, [3] and [4] have shown an overview on this active educational data mining area. Although of an applied field of the data mining research arena, educational data mining encountered many challenges stemming from the peculiar features of the educational application domain.

Many research works on educational data mining have recently been proposed in [3], [5], [6], [7], [8], [9]. These works used educational data gathered from many different education levels: [6] at the pre-university level, [3], [5], [7], [9] at the undergraduate level, and [8] at the graduate level. Moreover, they employed a diversity of data analysis and mining techniques, for example: [3], [6], [8] used statistical analysis and [5], [7], [8], [9] classification. Most of them used a data set gathered through a web-based learning system for a not very long period of time.

III. Proposed Ideas

In order to put forth the concepts which were briefly described above, we will consider a deemed university which has institutions spread across various geographic locations. The institution at each location can have multiple departments under them. Each department can further contain several branches. Each branch can have multiple years (4 years in case of engineering course). Each year will contain a fixed number of subjects. Figure 1 presents a visual explanation of the hierarchical breakdown.

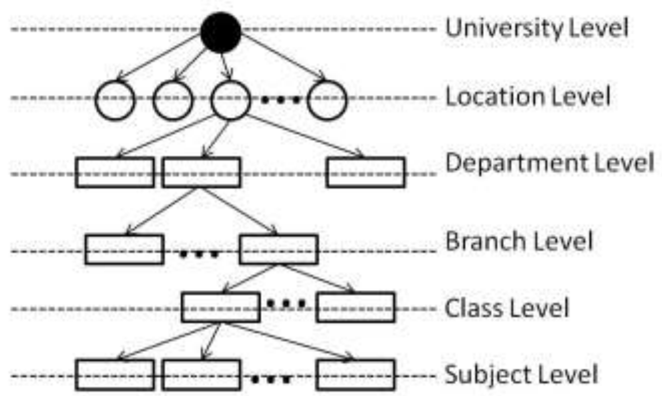

Fig. 1. A simple hierarchical breakdown of a university

The information which is mined out of the data collected from students' marks can be classified across various facts and dimentions. A fact table stores quantitative information for analysis and is often denormalized. A fact table works with dimension tables. A dimension is a collection of reference information about a measurable event / fact. A fact table holds the data to be analyzed, and a dimension table stores data about the ways in which the data in the fact table can be analyzed. If we consider the hierarchy shown in figure 1, we can start at the root level (university level) and perform a drill-down operation which will take you down to the location level. You may want to perform this drill-down if you want to see the individual performance of each location irrespective of the other locations. If you want to compare the performance of the various locations, you would perform a roll-up operation from the location level up to the university level.

Since exams are conducted every semester, and on an average a student answers 6 papers per semester, the amount of data generated in the form of exam results is enormous. Data mining methodologies can be applied to extract useful information from this pool of data. This data can then be pre-sented to the educators, recruitment managers and the students themselves. The information extracted can be presented baased on the category of people who want to view it, and they can analyze the data to come to useful conclusions.

Let us consider the example of a networking firm recruiting manager who is looking at hiring fourth year students for a networking job profile. A fourth year student would have studied and answered several networking subjects in the du-ration of his course. Since the recruiter would be interested in students who have excelled in networking subjects, we can mine out students who have qualifying scores in networking subjects, irrespective of their score in other non-networknig related subjects. This information can be presented in the form of a graph or a cube to the recruiting manager.

In this scenario, we can consider the students' marks as the fact, and we will consider only four dimensions to curb complexity. The four complexities are as follows:

- Location: This dimension specifies which location of the university he is looknig at.
- Student: This dimension will enclose the student's name and registration information which can uniquely identify him.
- Calender: This dimension will describe the time period that the recruiter is considering, for example, he could consider the last four semesters or he could consider the previous three years.
- Domain: This dimension specifies the area of specialization that the recruiter is looking at. The domain will be composed of several relevant subjects.

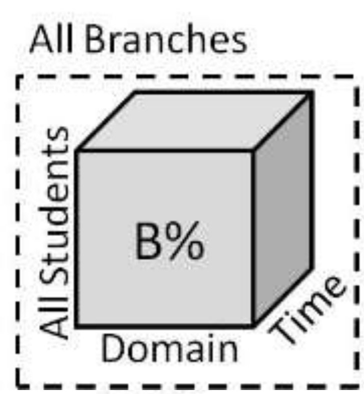

Fig. 2. Dimension - Fact Cube Representation

Consider figure 2 which represents the fact (marks obtained by students) across the various dimensions. B% indicates the aggregate marks (fact) obtained by all the students (student dimension) for a particular period of time, for example two years (calendar dimension), in a particular domain (domain dimension), for example Networking.

Now if the recruiter wants to see the aggregate marks obtained by each individual branch, he can perform a drill-down operation on the cube in figure 2. In this figure, "Branch ABC" and "Branch XYZ" denote the two branches (for example), "S" stands for "student", "NP" stands for "Network Programming Subjects", "11-12" indicates the academic year "2011-2012", "C%" and "D%" denote the percentage of marks obtained by the students in the respective branches.

With this drill-down operation, the recruiter now has information about the marks obtained by students in Network Programming Subjects in the individual locations.

Let us consider the information available for one location (Figure 4). If we perform a drill-down operation on the Students dimension on this cube (D1), we can get the marks obtained by each individual student as shown in the figure. From this point, we can perform drill-down operations on multiple dimensions. If we perform a drill-down operation on the Domain dimension (D2), it will display the marks obtained by the students in the individual subjects. If we perform a drill-down operation on the Calendar dimension (D3), it will display the marks obtained by the students across the years. If we perform a drill-down operation from these points (D4 or D5), we can see the marks obtained by each individual student in each subject across the various years.

From any of these points, we can perform a roll-up operation to go up a level. This will enable the user to drill into a different dimension to view a different perspective of the data.

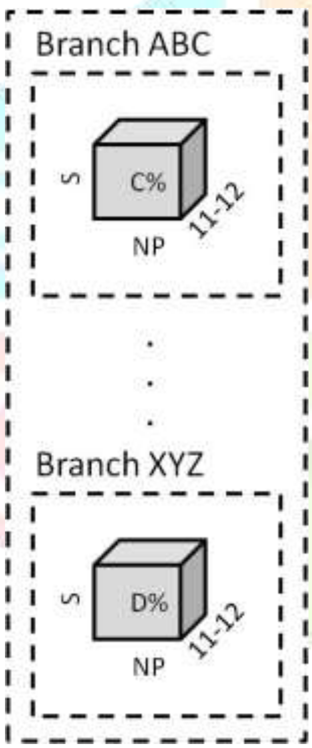

Fig. 3.

IV. Conclusion

This paper demonstrates the application of data mining in the analysis of educational performance across various facts and dimensions. With this approach, it is very convenient for different users to use the mined data in different forms. The above description shows how data can be used by industry recruiters. The same mined data can be used by educators to analyze the student and faculty performance. It can also be used to analyze educational trends and performance across the years.

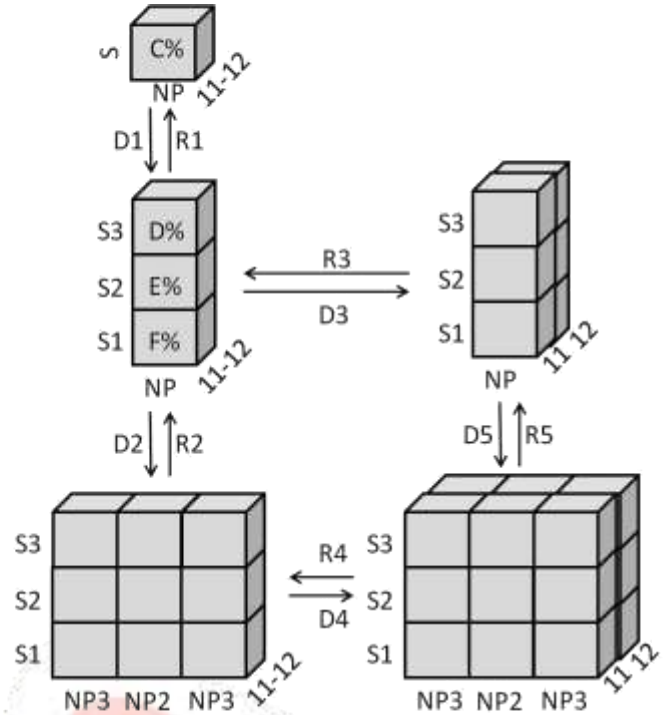

Fig. 4.

The advantage of this approach is that you dont need to execute data mining algorithms for each query that a user makes. Once the data has been mined, it can be saved and presented to the user depending on the way the user queries the data. This eliminates the need to execute high complexity algorithms for every request that is made.

The data can be presented to the user through a good Graphical User Interface, for example, through graphs and cubes. The drill-down and roll-up operations can be done using programming language libraries.

V. References

1) Vo Thi Ngoc Chau and Nguyen Hua Phung, ”A Knowledge-Driven Educational Decision Support System”, 2012 IEEE 3rd International Conference on.
2) Behrouz, M and Deborah, A and Gerd, K and William, F, ”Predicting student performance: An application of data mining methods with an educational web-based system”, 2003, 33rd ASEE/IEEE Frontiers in Education Conference.
3) J. E. Beck, ”Difficulties in Inferring Student Knowledge from Observations (and Why You Should Care)”, Edu-cational Data Mining, 2007.
4) P. R. Cohen and C. R. Beal, ”Temporal Data Mining for Educational Applications”, Int. J. Software and Informatics, vol. 3, pp. 29-44, 2009.
5) G. W. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers, ”Predicting Students Drop Out: A Case Study”, Educa-tional Data Mining, 2009.

6) Y. Gong, D. Rai, J. E. Beck, and N. T. Heffernan, "Does Self-Discipline Impact Students Knowledge and Learning", Educational Data Mining, 2009.
7) D. Kanellopoulos and S. Kotsiantis, "Towards an Ontology-based System for Intelligent Prediction of Stu-dent Dropouts in Distance Education", Int. J. Manage-ment in Education, vol. 2, pp. 172-194, 2008.
8) M. Paliwal and U. A. Kumar, "A Study of Academic Performance of Business School Graduates Using Neu-ral Network and Statistical Techniques", Expert Systems with Applications, vol. 36, pp. 7865-7872, 2009.
9) O. Taylan and B. Karagozoglu, "An Adaptive Neuro-Fuzzy Model for Prediction of Students Academic Per-formance", Computers & Industrial Engineering, vol. 57, pp. 732-741, 2009.