

Visualizing Educational Datamining Patterns

S. Gama¹ and D. Gonçalves¹

¹INESC-ID & Instituto Superior Técnico, Universidade de Lisboa

Abstract

Providing the educational community with the means to analyze the results of educational processes may help correct problematic situations, leading to a more successful education. Applying Data Mining to educational information provides the tools to analyze such data. However, results comprehend an extensive set of behaviors that consist of complex symbolic patterns. Visualization, due to its potential to display large quantities of data, may overcome this limitation. Hence, using the results of educational data mining that had been applied to information on courses in a university program, we developed a visualization that allows the analysis of such patterns. User tests have shown that our visualization makes important patterns immediately perceivable and displays the most relevant patterns, proving its potential for visualizing educational patterns.

Categories and Subject Descriptors (according to ACM CCS):

H.5.2 [Computer Graphics]: Information interfaces and presentation—User Interfaces

1. Introduction

The number of students has grown significantly over the last decades [Gao10]. Concerning online education, the profusion of MOOC (Massive Open Online) courses contributed to a growth in the number of students. CMS (Course Management Systems) and LMS (Learning Management Systems) became popular as well, with a great impact on distance education [KRDK13]. With the growing number of students in both traditional and online education, a large set of data emerges from students' curricula. This information may be crucial to improve education processes. The application of data mining techniques to educational data is an emergent research topic. However, the resulting data normally consist of an extensive set of behaviors that are described as textual patterns which are usually difficult to understand due to their visual complexity. Providing the tools for allowing an easy and correct interpretation of the information is, thus, of utmost importance. In order to overcome this obstacle and obtaining effective data analysis and prospection, the user must be involved in the exploration process in a way that combines flexibility and general knowledge [Kei02]. Visualization has the potential to meet such requirements. In fact, visualizing the results of educational data mining techniques will make it possible to provide the educational community with meaningful information, highlighting problems that would otherwise remain unnoticed, allowing the community to find adequate solutions.

The potential of visualization for displaying large quantities of data while alleviating cognitive load [War12] has been explored for displaying educational data, regarding online [MD05, MB07] and traditional education [XGHW13, XKP09, TRd12, WR07]. However, given the nature of information provided by applying data mining techniques to educational information, the majority of such techniques show several shortcomings, failing to allow course comparison or pattern highlighting. As a result, we developed a visualization which represents interrelations among a courses in an academic program while attempting to address the aforementioned shortcomings.

This paper is organized as follows: the next section presents our visualization, followed by a brief description of our evaluation in which several conclusions are drawn.

2. Visualizing Educational Patterns

We used the result of sequential pattern mining that had been applied to data gathered for nine years on an undergraduate study program on computer science [Ant08]. The goal of sequential pattern mining, given a set of sequences and some user-specified minimum support threshold, is to discover the sequences that exist in at least σ sequences in the dataset [AS95]. Sequential pattern mining with three different support threshold values (50%, 25% and 20%) has been performed, resulting in a number of textual patterns.

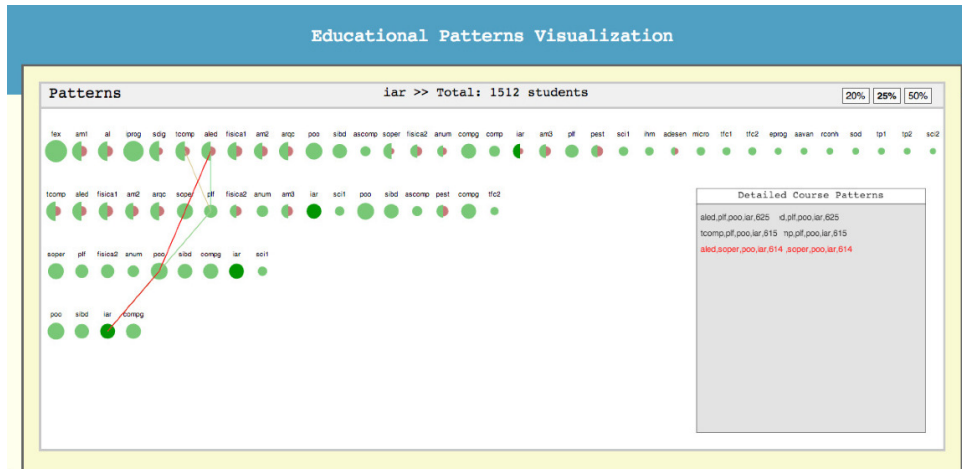


Figure 1: A multi-layered visualization for educational data mining patterns, revolving around course patterns' representation.

We created an interactive visualization (Figure 1). The main area consists of a multi-level visualization. Each level corresponds to a program's semester and represents its courses that are displayed as circles with size proportional to the number of students who completed or failed them. When there is information on failure, course circles are sub-divided into two semicircles displaying information on approval and failure through western conventional positive-negative color coding [War12]. Their size is proportional to course approval or failure numbers, making it possible to immediately comparing success and failure rates. Moving the mouse over a specific course circle highlights it (by assigning more saturation to its hue) and displays information on course relationships. The course is linked to all the courses with which it has any type of interrelation through line connectors. Connector thickness is proportional to the number of students who verify the pattern. Color is assigned to each pattern individually, in order to avoid visual confusion and allowing immediate line discrimination. We did not use fully saturated colors, in order to keep our visual artifacts from competing for the user's visual attention [War12]. When a course is selected, additional information on the total number of students is displayed on the additional course information panel, located at the top of the visualization (which also displays three buttons (20%, 25% and 50%) for the selection of different data mining support thresholds). The total number of patterns is shown at the upmost part of the detailed course patterns' panel and each pattern is listed below. Our visualization allows interactive course comparison. If we select one course, it is locked and its information is shown throughout the visualization (without being cleared when the mouse leaves the circle). Moving the mouse to another course will also display its information, making it possible to compare patterns. With this combination of mechanisms we can immediately perceive aspects such as the number of semesters

and the number of courses for each semester, as well as the most successful or unsuccessful courses, leading to the use of visualization for improving success rates.

3. Evaluation and Conclusions

We conducted a user study with 10 participants to evaluate the effectiveness of our solution to represent data mining educational patterns. Subjects were asked to perform five tasks, consisting of naming: (i) the number of semesters, (ii) the three most successful courses, (iii) the number of students who completed a given course, (iv) which courses had a relationship with a given course, and (v) the number of relationships that two different given courses had with other courses. Each task was performed three times, corresponding to the different support thresholds, regarding different courses. Time and errors were measured during task performance and a SUS (System Usability Scale) [Bro96] satisfaction questionnaire was handed at the end. Results suggested users took longer to perform tasks as complexity increases. We performed further statistical analysis to verify such impact, showing that complexity did not have a significant impact on either task performance time or number of errors, suggesting that the system scales well for large amounts of information. User satisfaction questionnaires yielded very good results, showing that our visualization displays the information in a way which provides the educational community with patterns that were not evident otherwise, providing information on how to act to promote further success.

Acknowledgments

This work was supported by national funds through FCT Fundação para a Ciência e a Tecnologia, under project Educare - PTDC/EIA-EIA/110058/2009 and INESC-ID multi-annual funding - PEst-OE/EEI/LA0021/2013.

References

- [Ant08] ANTUNES C.: Acquiring background knowledge for intelligent tutoring systems. In *Proceedings of the International Conference on Educational Data Mining* (2008), pp. 18–27. [1](#)
- [AS95] AGRAWAL R., SRIKANT R.: Mining sequential patterns. In *Proceedings of the IEEE International Conference on Data Engineering* (1995), pp. 3–14. [1](#)
- [Bro96] BROOKE J.: Sus: A quick and dirty usability scale, 1996. [2](#)
- [Gao10] GAO Y.: A study on mass higher education in the world-based on comparative perspectives. In *International Conference on Education and Management Technology (ICEMT)* (2010), pp. 528–530. [1](#)
- [Kei02] KEIM D.: Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics* 8, 1 (2002), 1–8. [1](#)
- [KRDK13] KAY J., REIMANN P., DIEBOLD E., KUMMERFELD B.: Moocs: So many learners, so much potential ... *IEEE Intelligent Systems* 28, 3 (2013), 70–77. [1](#)
- [MB07] MAZZA R., BOTTURI L.: Monitoring an online course with the gismo tool: A case study. *Journal of Interactive Learning Research* 18, 2 (April 2007), 251–265. [1](#)
- [MD05] MAZZA R., DIMITROVA V.: Generation of graphical representations of student tracking data in course management systems. In *Ninth International Conference on Information Visualisation, 2005. Proceedings* (2005), pp. 253–258. [1](#)
- [NI11] N. ILHINSKY J. S.: *Designing Data Visualizations*. O’Reilly, 2011.
- [Sau11] SAURO J.: *A Practical Guide to the System Usability Scale: Background, Benchmarks and Best Practices*. CreateSpace, 2011.
- [TRd12] TRIMM D., RHEINGANS P., DESJARDINS M.: Visualizing student histories using clustering and composition. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2809–2818. [1](#)
- [War12] WARE C.: *Information Visualization: Perception for Design*. Elsevier, 2012. [1](#), [2](#)
- [WR07] WORTMAN D., RHEINGANS P.: Visualizing trends in student performance across computer science courses. In *Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education* (2007), pp. 430–434. [1](#)
- [XGHW13] XIAOHUAN W., GUODONG Y., HUAN W., WEI H.: Visual exploration for time series data using multivariate analysis method. In *8th International Conference on Computer Science Education (ICCSE)* (2013), pp. 1189–1193. [1](#)
- [XKP09] XIAOYA G., KAN L., PING L.: Visual analysis of college students’ scores in english test. In *4th International Conference on Computer Science Education (ICCSE)* (2009), pp. 1816–1819. [1](#)