Visualizing Sequential Educational Datamining Patterns

Vilma Rodrigues Jordão, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
Sandra Gama, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal
Daniel Gonçalves, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal

ABSTRACT

The use of educational datamining techniques to elicit patterns in student behavior and learning outcomes can be a useful for the analysis of the effectiveness of teaching strategies and the coordination of study programs. However, results from those techniques are, often, large sets of symbolic patterns, numbering in the thousands, usually presented in text format. This makes them hard to understand which, coupled with the lack of an overall view, hinders a more comprehensive data analysis. The authors propose that information visualization techniques can be used to display relevant information in those patterns in effective ways, allowing decision makers to better insights about the reality at hand. They present a solution built upon two linked views, one based on node-link representations and another a multi-matrix representation. The complementarity of both visualization techniques allows the most important patterns to be immediately apparent, while at the same time permitting their interactive exploration in meaningful ways. The authors performed user tests proving their effectiveness.

KEYWORDS
Color Blending, Education, Educational Datamining, Human-Computer Interaction, Information Visualization, Sequential Patterns

INTRODUCTION

The number of students in both traditional and online teaching has grown considerably over the last decades. At the college level, there has been an increase from 8.5% (in 1970) to 24.7% (in 2006) in the percentage of people that have at least one degree (Gao, 2010). Online learning in particular has seen a huge boost given the emergence of Massive Open Online Courses (MOOCs), in which people from all over the world can enroll and participate. As a result, Course Management Systems (CMS) and Learning Management Systems (LMS) have become very popular and have now a large impact on distance learning (Kay, 2013).

To understand the effectiveness of a learning program or a particular course, a number of indicators must be considered, the not least important of which being student success. This in itself includes several factors such as drop-out rates and final grades but is, nevertheless, a simple indicator of how well things are going and can provide early warning signs for systemic problems that must be corrected. Also important are the inter-relations between courses in a same degree. Understanding them will give us insights into critical paths and bottlenecks in a curriculum. All of this can be used to fine tune curricula, leading to an increase in teaching quality, better learning outcomes and student satisfaction. Small student/course samples are not enough to elicit any meaningful, generalizable
information. However, given the aforementioned increase in numbers, coupled with the development of increasingly sophisticated LMS and the school’s own information systems, a wide and encompassing range of data is now within reach. Its timely analysis can be crucial for the improvement of the teaching-learning process.

Alas, the available data is seldom in a format that is amenable to analysis and exploration. Of real help for decision makers would be meaningful patterns that can be elicited from that data, rather than having to go through thousands of individual records. Only in that way can the overall picture be known and understood. The application of data mining techniques in this context is an emerging research field. It provides the means to analyze educational data, from student behaviors to teaching strategies and course coordination. For this domain, it takes the moniker of Educational Data Mining (Romero, 2010). EDM provides relevant patterns based on the available data and is a tool both for the study of what went on in the past and to make informed predictions for the future. However, the results from an EDM process usually consist of extensive sets of behaviors, described in the form of technical patterns that are represented textually or symbolically. While much more meaningful than the original data, the amount of patterns can still be daunting and not conducing to an encompassing analysis. Furthermore, the understanding of the patterns themselves often requires a reasonable knowledge of the underlying data mining algorithms and statistical models, which many analysts probably do not possess.

Globally, it is important to provide an overall view of all patterns as a coherent whole, allowing the identification of commonalities and differences. It is also necessary to interpret individual patterns and establish relations between them. Furthermore, a tool that allows this should involve the users and allow them to take advantage of their creativity, flexibility and domain-specific knowledge (Keim, 2002). One possible approach to this problem is to use Information Visualization. One of its aims is to help understand large amounts of data by leveraging on the capacity of the human visual system to discover trends, patterns and outliers (Heer, 2010). Furthermore, a well-designed visualization can effectively represent large amounts of data and alleviate the cognitive load associated with interpreting it (Ware 2004). A visualization of the results from EDM has the potential to provide the insights it needs.

In the context of the Educare project, the authors had access to sets of EDM patterns (Antunes, 2008). Their goal was the analysis of the interdependencies of college courses within an undergraduate degree. They are based on the number of students that pass or fail each course, and the sequence in which they do so. An example of a pattern (stated in natural language) is: 50% of all students that failed course A and passed course B on a given semester were not able to pass course C on the next semester they enrolled. This example illustrates the main challenge in creating a visualization for EDM patterns, besides their large numbers: their sequential nature. These patterns are structured as sequences of occurrences through time (in this case, school semesters). This sequential nature is crucial for their understanding and must be central in the visualization. However, most work in the visualization of data mining patterns focuses only on static patterns, representing a single instant of time.

In this paper the authors describe how a visualization, EduVis, was created, which allows the representation of sequential EDM patterns and their exploration using two complementary techniques: a multi-layer link-node visualization that makes dependencies evident and a multi-matrix-based visualization supported on heatmaps to highlight overall pass/fail trends. Both take advantage of color blending techniques to present more degrees of freedom without unduly overcomplicating the visualization in spatial terms. A set of user tests has validated their understandability, correctness and usefulness.

In the next section the related work in the area of the visualization of educational information will be described. Next, the nature of the patterns to visualize will be discussed. Following that, EduVis will be presented, its design rationale and underlying decisions, followed by its evaluation, supported on user studies, after which the authors will draw some conclusions, pointing to possible directions for future work.
3D Printing Meets Humanitarian Design Research: Creative Technologies in Remote Regions
Jennifer Loy, Peter Tatham, Ry Healey and Cassie L. Tapper (2016). Creative Technologies for Multidisciplinary Applications (pp. 54-75).
www.igi-global.com/chapter/3d-printing-meets-humanitarian-design-research/148564?camid=4v1a