International Journal of Creative Interfaces and Computer Graphics
Volume 9 • Issue 1 • January-June 2018

Visualizing Historical Patterns in Large Educational Datasets

Tiago Martins, Instituto Superior Técnico / University of Lisbon, Lisbon, Portugal
Daniel Gonçalves, INESC-ID / Instituto Superior Técnico / University of Lisbon, Lisbon, Portugal
Sandra Gama, INESC-ID / Instituto Superior Técnico / University of Lisbon, Lisbon, Portugal

ABSTRACT

With the increase in the number of students worldwide, it has become difficult for teachers to track their students or even for institutions themselves to identify anomalies in degrees and courses. The sheer amount of data makes such an analysis a daunting task. A possible solution to overcome this problem is the use of interactive information visualization. In this article, the authors developed a visualization that allows users to explore and analyze large datasets of academic performance data allowing the analysis and discovery of temporal evolution patterns for courses, degrees and professors. The authors applied the techniques to fourteen years of data for all students, courses, and degrees of a Portuguese engineering college. The system's usability and usefulness were tested, confirming its ability to allow analysts to efficiently and effectively understand patterns in the data.

KEYWORDS

Courses, Degrees, Educational Information, Historical Patterns, Information Visualization, Large datasets

INTRODUCTION

Education levels have been increasing across the world. This is true for traditional learning in brick-and-mortar schools and colleges, and even more so given the rise of new technologies in distance education such as Massive Online Open Courses (MOOC), supported by learning management systems (LMS) and course management systems (CMS). The number of students has, thus increased throughout the years and with it, the need for new tools to be able to understand of effective the learning experience is.

Platforms such as CMS and LMS allow the easy storage of records related to student grades, attendance in classes and approvals in a course. Analyzing this data may allow the detection of students with problems or even issues with the content and structure of the courses themselves (Baepler, 2010). The analysis of this data can not only allow a deeper understanding of eventual problems, but also helps alleviate the problems of a reality where students and professors are less knowledgeable of one another, either due to in-presence teaching massification or in on-line, distance learning settings, which make it difficult to perceive their problems and to track down their causes (May, 2011). This can lead to deeper reflections about the contents being taught and the development of strategies to minimize failure and lower dropout rates

Given the amount of data that these systems collect over multiple years, analyzing it may be a very difficult task and require high cognitive effort. A possible solution that allows the understanding and identification of relevant patterns in data is Information Visualization (Desai, 2014). By using

DOI: 10.4018/IJCICG.2018010103

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

interactive Information Visualization (InfoVis), one can explore different facets of the data and cope with the high complexity of very large datasets by using filtering and selection mechanisms. The use of derived measures that embody higher-level trends can provide precious different perspectives from the original information (Mazza, 2009).

We propose a solution where InfoVis is used to represent large datasets of academic performance data. A major challenge in this regard is scalability. While most previous approaches, as described in the Related Work section, deal with data from a single course or degree, we want to be able to address the problem at the school level. As such, we devised a set of minimalistic views which, while sacrificing some more advanced analysis, make up for that shortcoming with the possibility for long-term analysis of all degrees and courses of an entire university-level school. An additional hurdle, the heterogeneity of curricula between degrees and even inside the same degree through time (as curricular restructuring takes place) also contribute to make this a challenging domain.

As a proof of concept, we used the data from Instituto Superior Técnico (IST), the engineering school of the University of Lisbon, encompassing fourteen years and all undergraduate and MSc level degrees. To that end, we used official information sources made available through the school's information system, based on the FenixEdu platform that provides an API for certain data to be extracted.

Given the large dataset, studying the evolution of a degree or course over the years or discovering patterns and trends in the data could still become a time-consuming process without prior data processing (Ali, 2012). Thus, to make this data amenable to visualization, the entire dataset was pre-processed, cleaned and re-structured, as will be described below. The resulting dataset, with several thousand entries, was then successfully visualized by our solution, allowing interesting patterns to be found, as shown by usability and usefulness tests, performed with the help of IST's Statistics and Prospective Unit.

We, thus, developed a system that allows the analysis of the academic path of students and the evolution of degrees and courses over the years. This system may help professors, course coordinators and the School itself with assessing relevant issues that can be timely solved and improve the teaching-learning process.

In section two we explain the different works developed in the area, each exploring different mechanisms and academic data. The problems and strengths related to these works will also be discussed. In section three is presented the earlier phase of the implementation of our solution, which consists in the conception of the possible questions to which the visualization will answer and the description of the entire extraction process of the data. In the next section, we describe all the implementation phase, including the design of non-functional prototypes, the structure of the application and backend and frontend development. In the fifth section, the method used to evaluate the tool will be described. In section six the conclusion of the work is presented.

RELATED WORK

The increase in the number of students in traditional and online education has provided a large amount of educational data with the introduction of learning systems. Analyzing student behavior and the evolution of courses and degrees has become possible, even in the face of a deluge of information.

Considering the analysis of the academic data, such as the academic path and the evolution of the courses and degrees, there was a need to develop applications that allow the teachers and the coordinators to keep improving the students' performance and overcome the academic failure. However, due to the increasing number of students in online teaching and the amount of data generated by them, it is hard to keep track of the students' difficulties and to help them overcome these issues, based on unprocessed data.

Considering this scenario MOOCViz (Dernoncourt, 2013) was created, which provides visual techniques that may help the teachers to evaluate the performance of their students. The visual

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/visualizing-historical-patterns-in-large-educational-datasets/210549

Related Content

Data-Driven Maps of Art History
Doron Goldfarband Dieter Merkl (2019). *International Journal of Art, Culture, Design, and Technology (pp. 1-15).*
www.irma-international.org/article/data-driven-maps-of-art-history/233614

ConnectiCity: Real-Time Observation and Interaction for Cities Using Information Harvested from Social Networks
Salvatore Iaconesiand Oriana Persico (2012). *International Journal of Art, Culture and Design Technologies (pp. 14-29).*
www.irma-international.org/article/connecticity-real-time-observation-interaction/70392

Self-Objectification vs. African Conservative Features in the Selfies of Black African Women: A Study of Nigerian Social Media Users
Endong Floribert Patrick Calvain (2018). *Selfies as a Mode of Social Media and Work Space Research (pp. 103-130).*
www.irma-international.org/chapter/self-objectification-vs-african-conservative-features-in-the-selfies-of-black-african-women/191374

Drawings from Small Beginnings
Hans Dehlinger (2012). *Biologically-Inspired Computing for the Arts: Scientific Data through Graphics (pp. 255-266).*
www.irma-international.org/chapter/drawings-small-beginnings/65032

Preoperative and Intraoperative Spatial Reasoning Support with 3D Organ and Vascular Models: Derived from CT Data using VTK and IGSTK
João Fradinho Oliveira, José Luis Moyano-Cuevas, J. Blas Pagador, Hugo Capoteand Francisco Miguel Sánchez Margallo (2015). *International Journal of Creative Interfaces and Computer Graphics (pp. 56-82).*
www.irma-international.org/article/preoperative-and-intraoperative-spatial-reasoning-support-with-3d-organ-and-vascular-models/147172