Analyzing non-linear video usage in an introductory x-MOOC about basic linear algebra

Juan Triay, Julià Minguillón, Teresa Sancho-Vinuesa
Universitat Oberta de Catalunya,
Rambla Poblenou, 156. 08018 Barcelona, Spain
{jtriaypa, jminguillona, tsancho}@uoc.edu

Vanesa Daza
Universitat Pompeu Fabra,
C/ Tànger, 122-140. 08018 Barcelona, Spain
vanesa.daza@upf.edu

***Abstract*—Massive data collection in MOOCs provides insight about real course participation. The main aim of this work is to explore how MOOC students consume videos, and whether they follow a specific viewing pattern according with the characteristics of each scene in the storyboard. Using heatmaps as a visual analytics tool, results suggest that most participants do not alter the linearity of time, but in a few hot spots. This knowledge can provide course teachers with insightful information on how to detect possible bottlenecks and to improve video internal structure.**

1. Introduction

The use of interactive visualizations is a new way of providing all the stakeholders in education (i.e. learners, teachers and managers) with information to achieve their goals and take decisions while improving their reflection process. Visualization is becoming more than just a series of tools, technologies, and techniques for managing data sets, it is becoming a mass medium in itself, with many expressive possibilities, and is already starting to be one of the main tools and means for graphic user interface design and development [1], including visual analysis.

MOOCs are an ideal setting to analyze and visualize a huge amount of educational data, including video consumption [2]. UCATx, a platform based on Open EdX, offers a wide set of open courses, online, for free. It is a joint venture of the Catalan universities and the Catalan government (Generalitat de Catalunya), with the aim of opening their knowledge to the whole society and, at the same time, creating a framework for research and debate on online education and MOOCs. In this paper we analyze how UCATx students consume videos according to their intrinsic storyboard, in order to detect whether this consumption follows a linear pattern or not, detecting hot spots.

2. Methodology

The UCATx platform automatically captures students data of any course in .JSON format, related to their interaction with the course resources. We have developed several Python scripts in order to analyze data related to video consumption [3], detecting jumps within a video. A jump is when a participant moves the play bar forward or backward. These scripts capture three different items: YouTube video identifier, exact time when the participant begins to jump, and exact time when such jump finishes.

Using these data, we have developed different heatmaps to visualize jumps, using D3.js. We use two different color gradients, red for backward jumps and blue for forward jumps, as they have different meanings. A darker color means more jumps, using white in case of no jumps. A heatmap is a square matrix where y-axis represents the second where the student begins the jump and x-axis represents the second where the jump finishes. Therefore, the total number of cells is equal to the video length in seconds squared. Even for short videos, this temporal resolution is clearly excessive, with two major drawbacks: most of the heatmap is white (i.e. no jumps) and it takes a lot of time to render it. Therefore, we built series of heatmaps with different fixed temporal resolution, varying it from 2 to 20 seconds. Nevertheless, we observed that no optimal temporal resolution can be inferred from video length, jump position or jump length. Jumps depend on screen resolution and window size, which are not available.

In order to solve the abovementioned problems with fixed temporal resolution, we decide to generate the heatmaps according to the internal video structure (i.e. the storyboard), with no fixed temporal resolution. Each video was analyzed and then divided in different scenes, having different lengths. Therefore, each cell of the generated heatmap represents the scene where the student begins and finishes a jump. The number of jumps within each cell is normalized according to scene duration. The total number of cells is reasonably small (usually less than one hundred), thus simplifying heatmap rendering and reducing data sparseness. It also provides a clear representation with no too short jumps, reducing noise.

3. Results

Due to space constraints, we only show the four most viewed videos (V1-V4) data in Table 1 from an introductory algebra course. We only show the backward (BH) and forward (FH) jumps heatmaps for V1 (figures 1 and

TABLE 1. SUMMARY OF VIDEO CHARACTERISTICS.

Video	Jumps	Length (s)	Plays	Backward	Forward	Scenes
V1	1430	880	2834	728	702	15
V2	1393	541	2913	646	747	8
V3	1312	896	2509	737	575	12
V4	1260	504	2334	752	508	8

2, respectively), which contains information about integers divisibility and related properties.

Notice that most of the jumps (both backward and forward) happen within the same scene, and only a few cells out of the diagonal show some activity. Furthermore, jumps between cells are usually between consecutive cells. This may be a sign that video structure (i.e. the storyboard) is adequate and introduces concepts, examples and exercices in a logical order. Due to normalization, the highest concentration of jumps is not in the longest scenes, showing possible hot spots that require an interpretation related to the storyboard.

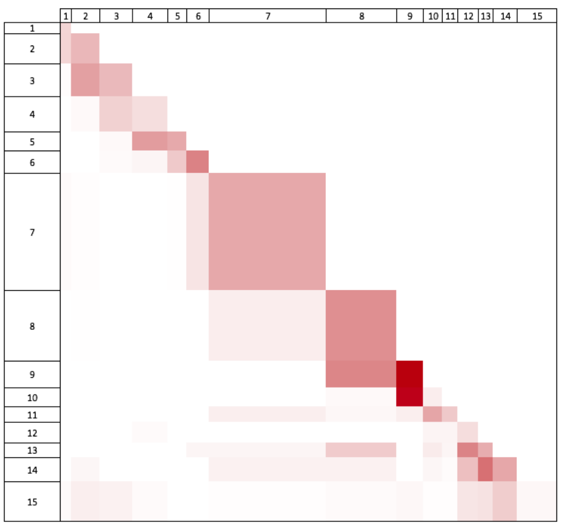

Figure 1. Heatmap of V1 backward jumps.

Figure 1 reveals interesting facts about this video. In general, back jumps mean reviewing previous content, in order to refresh a concept. Scene number 9 (an example about the properties of integer numbers) reveals that it may be too dense, as it generates a lot of jumps in a short period of time. On the other hand, Scene 10 introduces the concept of prime numbers, and it makes students to review the concepts previously introduced in scene 9. Therefore, an introductory example is crucial for the students in order to understand a specific concept, which also requires a clear and short explanation.

Regarding Figure 2, forward jumps have a completely different meaning, as students are skipping content, due to two main reasons: maybe they find it boring, thus advancing faster, or maybe they have already seen the video before and they are just skipping some parts. It is remarkable that the

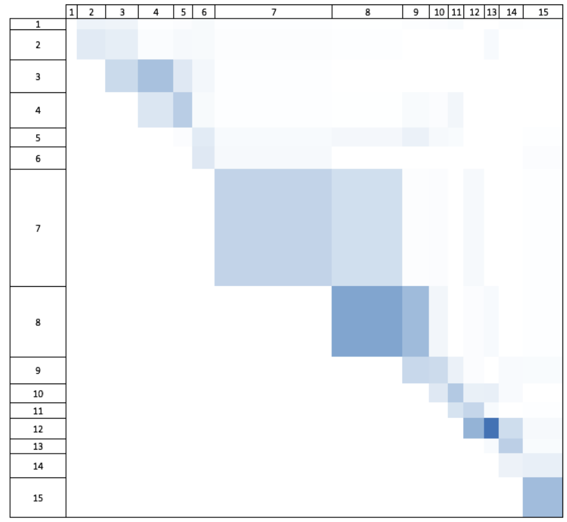

Figure 2. Heatmap of V1 forward jumps.

highest concentration of jumps is from scene 12 (a solved exercise) to scene 13 (a proposed one), probably because scene 11 is also a solved one, making scene 12 redundant.

4. Conclusions

Even with such a preliminary analysis it is possible to draw interesting conclusions about how course participants in an introductory algebra MOOC are consuming videos. Heatmaps based on video storyboard can be used to detect problematic scenes by analyzing backward and forward jumps. Heatmaps show that the majority of participants jump within the same scene. If not, jumps are performed to the previous or from the posterior scene. Each heatmap is a signature that can be analyzed to determine possible hot spots and bottlenecks.

Current and future research in this topic includes automatic extraction of video storyboard, fuzzy mechanisms for improving boundaries between consecutive scenes and comparing heatmaps to characterize (and compare) videos.

Acknowledgments

This work has been partially supported by Catalan government grant 2014 SGR 1271 (LAIKA).

References

[1] L. Manovich, "What is visualization?" *Visual Studies*, vol. 26, no. 1, pp. 36–49, 2011.

[2] N. Li, L. Kidzinski, P. Jermann, and P. Dillenbourg, "How do in-video interactions reflect perceived video difficulty?" in *Proceedings of the European MOOCs Stakeholder Summit 2015*, Mons, Belgium, 2015, pp. 112–121.

[3] J. Triay, J. Minguillón, T. Sancho-Vinuesa, and V. Daza, "Exploring the effectiveness of video viewing in an introductory x-mooc of algebra," in *Trends in Digital Education: Selected papers from EC-TEL 2015 Workshops CHANGEE, WAPLA, and HybridEd*, vol. 1599, Toledo, Spain, September 2015, pp. 63–70.