VisMOOC: Visualizing Video Clickstream Data from Massive Open Online Courses

Conglei Shi† Siwei Fu‡ Qing Chen† Huamin Qu§

ABSTRACT
Massive Open Online Courses (MOOCs) are becoming increasingly popular and have attracted much research attention. Analyzing clickstreams on MOOC videos poses a special analytical challenge but provides a good opportunity for understanding how students interact with course videos, which in turn can help instructors and educational analysts gain insights into online learning behavior. In this poster, we develop a visual analytical system, VisMOOC, to help instructors analyze the clickstream data. VisMOOC consists of three main views: the List View to list all course videos for analysts to select the video they are interested in; the Content-based View to show how each type of click actions change along the video timeline, which enables the most viewed sections to be observed and the most interesting patterns to be discovered; The Dashboard View shows the information of the clickstream data in different aspects, including the course information, the geographic distribution, the video temporal information, the video popularity, and the animation. Furthermore, case studies made by the instructors demonstrate the usefulness of VisMOOC and helped them gaining deep insights into learning behavior for MOOCs.

1 INTRODUCTION
Massive Open Online Courses (MOOCs) have attracted a lot of public attention over the past few years [4, 5]. The considerable amount of data generated from MOOCs offer a great opportunity for educational analysts to analyze the learning behavior [1]. Relevant data include student profiles, posts in the course forum, surveys, course videos and clickstreams of the course videos. Particularly, in the clickstreams, there are six types of clicks, namely, “play”, “pause”, “seek”, “stalled”, “ratechange”, and “error”.

A lot of statistical studies have been done to analyze the data from different aspects, providing valuable insights into learner behavior in MOOCs [1, 2, 6, 7]. Particularly, recent research shows that students who take online courses spend the majority of time watching lecture videos [1, 6]. Therefore, it is important for instructors to understand how learners behave when watching videos. For instance, they can revise the video accordingly to make it more comprehensive by better understanding of learning behavior.

Recently, a large-scale analysis of click streams for lecture videos has been reported [3]. This analysis provides insights into dropout behavior and reasons underlying the interaction peaks in videos. Also, it is the first work to study the click-level interactions in MOOC videos. However, when we interviewed the instructors of MOOCs, they said that there lacks tools for them to analyze learning behavior.

In this project, we collaborated with domain experts to iteratively design VisMOOC, a visual analytical system to help them understand online learning behavior. We used the log data of course videos and followed a user-centered process to develop the system.

To demonstrate the usefulness of our system, case studies are conducted about how experts used VisMOOC to explore the data and what they found. To the best of our knowledge, our study is the first to provide such a visual analytical system for domain experts to combine content-based analysis with video clickstream data of lecture videos.

2 VisMOOC DESIGN

The main interface of VisMOOC consists of three coordinated views that show clickstream data in different aspects as well as at different levels of details. The List View shows the list of all course videos, and analysts can select the video they are interested in. The Dashboard View shows the information of the clickstream data in different aspects, including the course information, the geographic distribution, the video temporal information, the video popularity, and the animation. We also support multiple interactions such as filtering and selecting.

The Content-based View is the center part of our system, which provides an in-depth analysis of the clickstream along with the video content. In this view, two visualizations are used to encode different types of information. The event graph shows the distribution of events on a video. We construct second-by-second counts for six types of events and use a stacked graph to visualize them. We use the color channel to encode the event type, and the height is used to encode the number of events. The graph helps analyze how learners are engaged with the video content. The seek graph uses two parallel axes to encode the starting position and the ending position of seeks. A line is drawn between two axes to connect the starting and ending positions together for each seek event. We use different colors to encode seek events happened on first watching (blue) or reviewing (orange). The upper part of the seek graph indicates forward seeks while the lower part represents backward seeks. We align the video with three visualizations using a highlighted line to help connect the video content and detailed clickstream information together for better analysis.

†e-mail: clshi@cse.ust.hk
‡e-mail: sfuaa@connect.ust.hk
§e-mail: jane.qing.chen@gmail.com
§e-mail: huamin@cse.ust.hk
All the authors are from CSE Department, the Hong Kong University of Science and Technology

Together, these views form a complete system that allows analysts to analyze the clickstream data.

3 Findings

To demonstrate the usefulness and effectiveness of our system, we conduct case studies together with five experts. They can freely explore the clickstream data by using VisMOOC with their own computers. We select the major findings below.

Finding I: We selected two user groups from the U.S. and China. The event graph clearly shows that the percentage of seek events in the U.S is much larger than the one in China and this was not a single case: we explored more videos and found that the clickstream data of all the videos followed the same pattern. The experts tried to explain the phenomenon. One possible reason is that, from their own experience in face-to-face education, more Chinese students prefer taking notes. Thus, when watching MOOC videos, Chinese learners may prefer pausing the video, taking notes, and playing it again. This can explain the lower percentage of seek events of Chinese learners.

Finding II: By exploring the data, the experts found that for different in-video questions, although the event graphs show a similar pattern, in the seek graphs, we can clearly observe the differences. For instance, Figure 3 shows the differences between seek graphs of two videos with in-video questions. For the video on the left, there is a considerably large percentage of backward seeks around the question while there are less backward seeks around the left, there is a considerably large percentage of backward seeks.

Figure 4: Two Content-based Views for the same video shown in Figure 1 but with different time periods. a) the clickstream data from the first week when the video was released; b) the clickstream data from the week when the related assignment was released.

Finding III: When the experts saw the Content-based View in Figure 1, they pointed out a strange pattern in the seek graph. There are two positions with dense backward seek events, however, the earlier position ($P_1$) is mostly filled with orange seeks that happened when learners reviewed the video, and the later one ($P_2$) is dominated by blue seeks that happened when learners first watched the video. From the event graph, both positions correspond to a peak. By examining the video content at position ($P_1$), the instructors of this course figured out that the video content relates to the assignment as well as the final exam. Thus, when we selected only the clickstream data before the assignment and the exam (Figure 4(a)), the first peak of re-watched seek events disappeared. To further confirm whether the assignment or the exam led to the phenomenon, the instructors selected the assignment release day and the exam day separately, and finally confirmed that this pattern was triggered by the assignment (Figure 4(b)).

4 Conclusion and Future Work

In this paper, we have presented VisMOOC, a visual analytical system to help instructors and educational analysts understand online video watching behavior using clickstream data from Coursera. In the future, our work can be extended in two directions. The first direction is to extend VisMOOC by analyzing forum data and grading data together with clickstream data. Such conjoint analysis could further help experts understand the learning behavior. For instance, the grading information can provide the ground truth about how well the learners perform, which would offer a new opportunity to identify different learning patterns between learners who get higher grades and those with lower grades. The other direction is to generalize this system for analyzing general online videos, which would benefit online advertising and video making.

References