MOOC Video Interaction Patterns: What Do They Tell Us?

Nan Li(✉), Łukasz Kidziński, Patrick Jermann, and Pierre Dillenbourg

École polytechnique fédérale de Lausanne, Lausanne, Switzerland
{nan.li,lukasz.kidzinski,patrick.jermann,pierre.dillenbourg}@epfl.ch

Abstract. For MOOC learners, lecture video viewing is the central learning activity. This paper reports a large-scale analysis of in-video interactions. We categorize the video behaviors into patterns by employing a clustering methodology, based on the available types of interactions, namely, pausing, forward and backward seeking and speed changing. We focus on how learners view MOOC videos with these interaction patterns, especially on exploring the relationship between video interaction and perceived video difficulty, video revisiting behaviors and student performance. Our findings provide insights for improving the MOOC learning experiences.

Keywords: MOOC · Online education · Video analysis · Video interaction · Clustering analysis

1 Introduction

Online education, with the recent revolution of the Massive Open Online Courses (MOOCs), is beginning to show itself as a disruptive innovation. MOOCs, as their current forms, typically replicate traditional classroom pedagogy online, featuring with video lectures, quizzes, tutorials, discussion forums and wikis. As recent studies [2,12] show, video lecture viewing is the central activity in MOOC learning, but little research work has yet endeavored to analyze click-level video interactions. For MOOC instructors, it is of great importance to understand how the students interact with lectures videos as well as how they perceive them.

The availability of large-scale MOOC data has made deeper video interaction analysis possible. As thousands of students interact with the same set of videos, we can plausibly categorize video interactions into groups of similar patterns, hereafter referred to as **video interaction patterns**. These patterns describe how the students typically use MOOC videos to achieve their learning goals. This paper attempts to categorize MOOC video interaction patterns by employing a clustering-based methodology, based on the available types of interactions, namely, pausing, forward and backward seeking and speed changing. Our hypothesis is that students adapt their video interactions to the video difficulties, their personal capability and learning strategies. So our analysis in this paper is further extended to explore the potential associations between video

G. Conole et al. (Eds.): EC-TEL 2015, LNCS 9307, pp. 197–210, 2015.
DOI: 10.1007/978-3-319-24258-3_15

interaction patterns and the following aspects: (1) perceived video difficulty (2) video revisiting behaviors (3) students' performance.

In this paper we categorize the video interaction patterns, but more significant contributions lie in the application of these patterns to reveal the relationship between video interactions and other factors, in order to address educational and design issues in MOOC. We summarize our contributions below:

- Analyzed the associations between video interaction and subjective video difficulty, and found video interaction patterns reflect perceived video difficulty.
- Identified the video interaction patterns that are linked to higher probability of video revisiting.
- Investigated the differences in video interaction patterns between weak and strong students, finding the weak students pause significantly more, and mostly during the presentation of code examples.
- Provided data-driven insights on improving the MOOC video interface

2 Related Work

2.1 Video Interaction Analysis

Videos players typically offer a limited types of interactions, each of which is associated with a time span. The sequential execution of the actions entail the Markov model a popular approach for video analysis. In early research, such analyses mostly aimed at evaluating the quality of service issues [3,10,13]. Research that attempted to model video click behaviors came to light since Branch et al. found that video interaction behaviors, in terms of the time spent on each viewing mode (i.e. play, pause, fast-forward, fast-rewind) can be modeled with lognormal distributions [1]. They also proposed a first-order Markov chain model for modeling different types of actions. Later Syeda-Mahmood et al. studied subjective video browsing states with a Hidden Markov approach [14], with the goal of generating video previews that best represents interesting video segments. All of the above studies were conducted in the time when the control menu of video players were restricted to only continuous interactions, lacking discontinuous interactions that are common in today's video player controls, such as seeking forward/backward, which allow jumping between different time positions.

Research on clustering video interaction behaviors also started before the MOOC era. In the early 2000s, Mongy et al. [11] proposed a method to apply K-means clustering with the Kulbach-Leibler distance between the state-transition matrices, but little is discussed about the data collection, the validation of the results and the scalability of the approach.

2.2 MOOC Analysis

With the rapidly popularized MOOCs, researchers have now enormous data for learning student's behaviors. Research on MOOC video interactions are usually centered on analyzing lecture-to-lecture navigation strategies [5], predicting dropout [6,15] and in-video dropouts [7]. Among these work, only [7,15] contain click-level video analysis. [7] emphasizes on the video interaction peaks while ignoring other silent interactions such as speed changing. [15] analyzed the video interactions sequences with n-gram analysis, which aimed at predicting dropouts. As an intermediate result, the information processing index computed from the click sequences serve for identifying and interpreting different video behaviors. The main problem is that the analysis did not consider the duration of each action. Furthermore,the interpretations of the click events are mostly based on "common sense knowledge" with little empirical evidence.

Existing MOOC literature with clustering analysis mainly aim at categorizing MOOC students' engagement [8], where the authors adopted unsupervised K-means method based on the students' longitudinal online learning activities. As far as we are concerned, clustering video behaviors with click-level interactions remain gap in MOOC research.

3 Video Interaction Dataset

We analyze two undergraduate MOOCs offered in Cousera: "Reactive Programming (RP)" and "Digital Signal Processing (DSP)". The former covers advanced topics in programing and the latter is a foundation course in Electrical Engineering. Table 1 summarizes descriptive information of these two MOOCs. Both courses have similar presentation styles, i.e. the professors present the lecture with PowerPoint slides, with a digital pen as both pointer and annotator.

Table 1. Overview of the two MOOCs in our dataset.

Subject	Week	Videos	Length	Quiz	Active	Passed	Sessions	Events
RP	7	36	18:50	6	22,794	5,276	470,994	4,001,992
DSP	10	58	16:20	16	9,086	263	117,959	1,138,558

The RP course attracted much more active students (those who at least watched a video) than the DSP course. There were no mid-term or final exams in both courses. Instead, assessments of students were made with weekly quizzes. Students' grades were computed as the sum of their best quiz scores of all trials in each week. The RP course allowed an unlimited number of quiz submissions, while the DSP course permitted five submission per quiz at maximum. The differences in grading schemes perhaps led to higher pass rate in the RP course (23.15 %), compared to only 2.89 % in the DSP course.

3.1 Data Wrangling Pipeline

In our data wrangling process, we first reconstruct the watching histories of each user from the raw video interaction logs. For each unique video in our dataset, user-based watching histories are created by arranging the events in chronological order. Revisited and first-time video sessions are separated. The next step is to aggregate the events in each watching segments into video features, which will be explained in detail in Sect. 4.

The processed video events include pauses, seeks and speed changes. The number of events shown in Table 1 includes the video events in all video sessions. Coursera does not only generate pauses when a user clicks the pause button. Automatic pauses are generated when an in-video quiz pops up or when the video is played to the end. Such automatic pausing events are removed for the analysis in this paper. In addition, students usually watch the lecture videos in uncontrolled environments, so the pauses are found to last for a maximum of several days. We removed the pause events that have a duration of more than 10 min, which are rather "breaks" than "pauses". In addition, a seek event is created when the user clicks or scrubs the playhead to a new position on the time bar. When scrubbing interactions occur, the logging system automatically generates a number of intermediate seeking events. Following the approach in [15], the seek events that are within 1 s interval are grouped as a single seeking event.

4 Video Interaction Clustering

Our first goal is to cluster video sessions into categories that characterize the video interactions. We computed a set of video features for each interaction type to characterize both the frequency and time dimension of the interactions. The features used for clustering are listed in Table 2.

Table 2. Video features used for clustering

1. Number of pauses (NP)	5. Number of backward seeks (NB)
2. Median duration of pauses (MP)	6. Replayed video length (RL)
3. Number of forward seeks (NF)	7. Average video speed (AS)
4. Proportion of skipped video content (SR)	8. Effective video speed change (SC)

Most of the features in the list are self-explanatory, but we will explain some of them in details. A video session may contain pauses of different durations. We use median statistic for pauses because it is more robust compared to the mean and sum statistics for the highly skewed, long-tail distribution of the pause duration data. As for the video speed, Coursera video player offers 7 levels of speed ranging from 0.75 to 2.0 with a stepwise change of 0.25 and the video player inherits the video speed from the users' previous sessions. That is, a video may contain no speed changing events, even though it is played at a speed other than 1.0. The *average video speed* feature refers to the weighted arithmetic mean of the video speeds at all video seconds. Since the videos may start at different speeds, the *average speed* feature alone may not always tell if the changes have happened during the video session. In combination with this feature, we also use *effective video speed change*, which is the average amount of speed change during the video session. This feature is computed by subtracting the starting video speed from the *average video speed*. We don't use the frequency of speed change events because it is inflated, i.e. the speed always has to be changed in multiple steps of 0.25. The datasets contain a large number of video sessions with **in-video dropout**, i.e. the student left the video before they reached the end. Such behaviors are different from skipping with forward seeks, because the user never reached later part of the video. When we compute the *proportion of skipped video content*, we only consider the proportion that is skipped by forward seeks. While discarding the video sessions that do not reach the very end sounds over-correct, we group all the video sessions that did not reach the last 10 % into the "in-video dropout" category, and our unsupervised clustering will be performed only on the remaining "complete" video sessions.

4.1 Clustering Pipeline

The datasets contain a large number of video sessions with no video interaction events (e.g. 17 % for the RP). These video sessions are filtered out and clustering is performed on the remaining dataset of the two MOOCs independently with the 8 video features presented before. After preprocessing with PCA dimension reduction, we obtain 6 new uni-variance variables which account for 90 % of the original variance.

The goal of clustering is to obtain a minimal number of interpretable clusters explaining user behaviors. For clustering, we use Neural Gas, a neural network-based convex clustering algorithm which is a robustly converging alternative to the k-means. The Simple Structure Index (SSI) [4] is used as a criterion for selecting the optimal number of clusters, since this index is known to multiplicatively combine several elements which influence the interpretability of a partition solution. For the RP course, we vary the number of clusters from 3 to 15, and find that 9 clusters maximize the SSI value (0.356 in [0,1] scale), as compared to the minimum value of 0.1 with 5 clusters. We then partition 9 video interaction clusters for the dataset. Similarly, 9 clusters are obtained for the DSP dataset as well.

4.2 Video Interaction Patterns

The centers of the 9 clusters for the RP dataset are shown in Table 3, and the results for the DSP course are analogous. The full names for the abbreviated feature names can be found in Table 2. We label each cluster with an intuitive name according to the corresponding dominating features (marked as bold) in the table. For example, the *LongPause(LP)* video sessions have an average median duration of pauses (MP) of 284.96 s. Note that the average number of pauses (NP) for this cluster is small (1.71). So it actually represents video sessions with infrequent long pauses.

Table 3. Cluster centers for the RP dataset

Pattern	Proportion	NP	MP	NF	NB	SR	RL	AS	SC
Replay (RP)	3 %	4.73	62.58	5.86	12.84	0.05	**531.44**	1.10	−0.00
HighSpeed (HS)	10 %	1.17	23.19	1.18	0.95	0.10	27.14	**1.66**	−0.01
SpeedUp (SU)	3 %	1.38	27.16	1.66	1.04	0.09	25.13	1.53	**0.39**
SkimSkip (SS)	4 %	1.00	30.73	**21.70**	4.94	**0.75**	17.46	1.14	0.00
Inactive (IA)	38 %	1.93	39.05	0.71	1.28	0.03	36.65	1.05	−0.00
FrequentPause (FP)	4 %	**13.39**	40.58	2.87	5.13	0.05	109.37	1.08	−0.00
JumpSkip (JS)	13 %	0.45	11.62	**5.38**	1.10	**0.71**	9.40	1.06	0.00
LongPause (LP)	6$	1.71	**284.96**	1.34	1.26	0.08	44.62	1.07	0.00
SpeedDown (SD)	1 %	2.13	42.93	1.61	1.73	0.08	44.42	1.24	−**0.58**

While Table 3 only presents the centroids of the clusters, the distributions of these features are illustrated as boxplot in Fig. 1, in terms of standard scores (z-scores) of the variables. 50 percent of the data points are enclosed in the boxes. The upper and lower whiskers extend from the hinge to the highest or lowest value that is within 1.5 interquartile range of the hinge. Data beyond the end of the whiskers are considered as outliers and are not shown in the figure. The multidimensionality, continuity, skewed distribution, and inter-correlation natures of the features imply that clear separations are unlikely to be found based on the current feature sets, and this explains why the maximum SSI (0.356) of the partition solutions is relatively small. Nevertheless, the dominating features in each cluster are still prominent, as shown in Fig. 1.

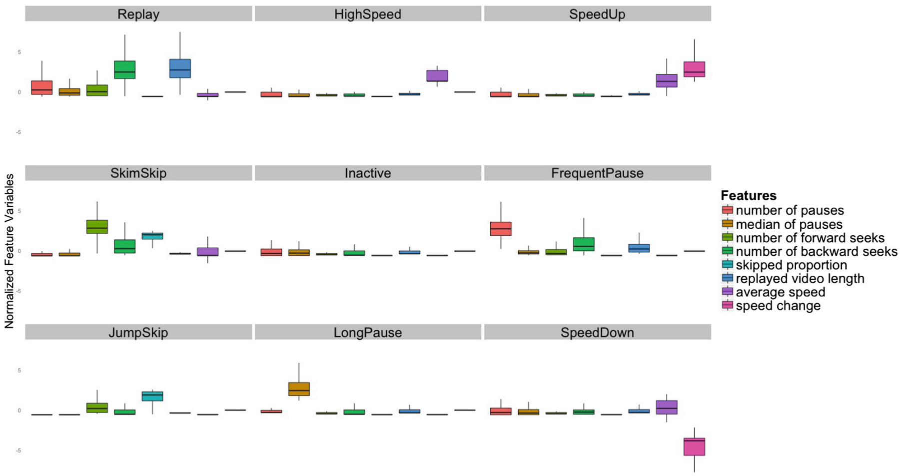

Fig. 1. Cluster data distributions for the RP dataset

In addition to the presented 9 patterns, we have (17 %) of the data removed before the clustering because they contain no video events. These sessions are labeled as *Passive (PS)*. It should be noted that most video sessions contain few video events, i.e. *PS*, *IA* and *HS* account for 65 % of the dataset, indicating a small number of video interactions satisfy the students' need for most of the time. On the other hand it also implies the adoption of rarer patterns may reflect certain changes in the students' learning state. We will discuss them in the upcoming sections.

5 Perceived Video Difficulty

Perceived video difficulty, i.e. how easy or difficult it is for a student to understand the content of a video, is a subjective measure of learning experiences. This measure reflects the students' cognitive states while watching lecture videos. We hypothesize that MOOC students may adapt their video interaction strategy to the video difficulty. Thus, our research question is: ***How do the different video interaction patterns reflect different levels of perceived video difficulty?***. In our recent work [9], we studied each type of interaction separately, but in this section we will examine the relationship between perceived video difficulty and video interaction patterns. The analysis may provide insights for MOOC practitioners to detect when a student may encounter difficulties.

5.1 Method

The perceived video difficulty is subjective and cannot be measured directly from video interactions, so we used in-video surveys placed at the end of each video during the enactment of the two courses to assess the subjective video difficulty. Only one question was asked: *How easy was it for you to understand the content of this video?* These surveys are posteriori evaluations that were typically answered by the learners right after they finished watching the video content, providing ground-truth knowledge of their situational perceived difficulty. The surveys were not graded, so the students participated voluntarily. The responses were then coded with integer values from 1 to 5 to represent the difficulty ratings from "Very Easy" to "Very Difficult". Students may watch the videos multiple times and leave more than 1 ratings for the same video. In the analysis of this section, we will only focus on the ratings of the first watching sessions. The response rate for the RP course is 49.1 % with an average difficulty of 2.699. For the DSP course, the rate is 32.8 % with an average difficulty of 2.594.

Since in our datasets the same users are measured multiple times, we used mixed-effect model by grouping the users as random effects to estimate the mean video difficulty. Mixed model are known to be robust to missing values and unbalanced groups, and Least-square means mimic the main-effects means but are adjusted for imbalanced group sizes.

5.2 Result

Mixed-model ANOVA shows significant effects of video interaction patterns on perceived video difficulty (RP: $F(9,124964) = 313$, $p<0.0001$; DSP: $F(9, 17505) = 24$, $p<0.0001$). We plot the Least-square mean difficulty with confidence interval in Fig. 2. The colored labels underneath the name of each pattern along the x-axis depict the number of video sessions with difficulty ratings and the total number of video sessions belonging to the corresponding pattern. The two numbers are separated with a slash sign "/".

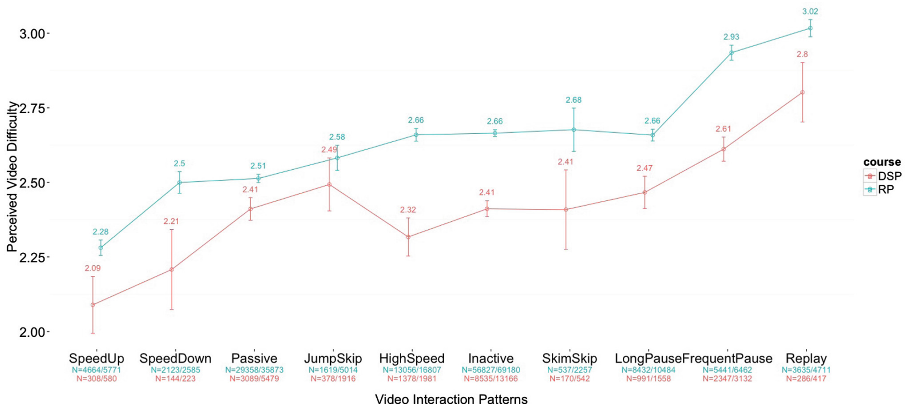

Fig. 2. Video interaction patterns and perceived video difficulty

Figure 2 shows that relative least-square mean differences in perceived video difficulty of different video interaction patterns are more or less consistent across two courses (with a systematic difference attributed to the course intrinsics), though the clusters are generated independently. Therefore, clusters generated from another course are likely to follow a similar trend.

The *Replay(RP)* and *FrequentPause(FP)* patterns reflect significantly higher video difficulty. They are employed as strategies to cope with difficult videos. Students may use the former to clarify doubts within the videos by rehearing the explanation, while the latter may be used when the explanations cannot be found within the video (requiring external resources) or the verbal/visual explanations are too fast to be processed (requiring buffer time)

The *SpeedUp(SU)* pattern reflects significantly lower video difficulty compared to the other patterns, indicating explicitly increasing the speed during video playback are commonly used by the students to adapt to easy videos. This way they can quickly grasp the gist of the video without skipping content. However, video sessions with *HighSpeed(HS)* pattern, i.e. the sessions that were started with high speeds inherited from previous sessions do not show significantly higher difficulty, compared to those with *Inactive(IA)* pattern. This suggests that the correlation effect of inherited speed on perceived video difficulty is weak, and may reflect personal preference rather than video difficulty.

Note that video sessions with large skipped content or with inherited higher speeds are not necessarily associated with easy videos. On the other hand, video sessions containing long pauses do not reflect significantly higher difficulty compared to *Passive* and *Inactive* patterns.

6 Video Revisiting Behaviors

Compared to traditional classroom lectures, MOOC videos are permanently preserved online, which makes revisiting certain videos a common practice. Students may revisit an older video for checking concepts while watching new videos or doing homework. Kim et al. find out [7] that first-watching sessions are more sequential while the revisiting sessions are more selective, i.e. the students selectively navigate the video into specific parts. We will inspect another perspective of video revisiting behaviors by asking ***With which first-time video interaction patterns are the videos more likely to be revisited?***. The analysis in this section may provide insights to MOOC designers for redesigning the user interface for improving the video revisiting experiences.

To start with, we first take a global overview of video revisiting behaviors on all the video sessions, including those "in-video dropout" video sessions, which are excluded for the analysis in this paper so far. We start by comparing the video sessions that are completely viewed and those containing "in-video dropout", because we found "in-video dropout" is strongly related to the occurrences of revisiting video sessions. As Table 4 shows, around one fifth of the completed videos (the "Completed" column) are revisited later. In comparison, videos that contain in-video dropout (the "Dropped-out" column) in the first-time watching sessions are significantly more likely to be revisited (73.7 % for the RP and 59.3 % for the DSP), according to the reported Chi-squared statistics. Note that in this paper all the Chi-squared tests hereafter are conducted with frequency of occurrences, though in the table we use percentage for presentation.

If we compare the video interaction patterns for complete video sessions only, then the result is as shown in Table 5. In each cell the percentage represents the observed probability of revisiting after the first view with the corresponding patterns. The expected probability for a video to be revisited for RP and DSP courses are 20.1 % and 23.5 % respectively, under the null hypothesis that video revisiting is independent of the interaction patterns. Chi-Squared tests show that the chances of revisiting significantly depend on the first-time video interaction patterns. Post-hoc residual analysis further reveals which cells contribute most to the Chi-Squared value. This is expressed by the adjusted standardized residuals, as shown below the percentage values in each cell. Significant positive residuals at $\alpha = 0.05$ (adjusted standardized residuals that are more than 2) are highlighted in bold. These highlighted cells indicate the frequency of occurrences for the corresponding patterns are significantly overly observed with respect to the expected frequency.

Table 4. Proportion of video revisiting for complete and in-video dropout sessions

	RP		DSP	
	Completed	Dropped-out	Completed	Dropped-out
Revisiting	20.6 %	73.7 %	23.5 %	59.3 %
No Revisiting	79.4 %	26.3 %	76.5 %	40.7 %
	$\chi^2(1{,}220875) = 55805.1$, $p < .0001$		$\chi^2(1{,}38825) = 5114.1$, $p < .0001$	

Table 5. Proportion of video revisiting for complete sessions

		RP	HS	SU	SS	IA	FP	JS	LP	SD	PS
RP	Revisiting	35.7 %	17.2 %	15.0 %	25.6 %	21.6 %	26.1 %	21.6 %	21.1 %	21.3 %	20.6 %
		25.9	−11.3	−10.5	**5.96**	**8.93**	**11.1**	1.86	1.31	0.99	−16.8
	No Revisiting	64.3 %	82.8 %	85.0 %	74.4 %	78.4 %	73.9 %	78.4 %	78.9 %	78.7 %	79.4 %
		−25.9	**11.3**	**10.5**	−5.96	−8.93	−11.1	−1.86	−1.31	−0.99	**16.8**
		$\chi^2(9{,}156517) = 1293.7$, $p < .0001$									
DSP	Revisiting	41.5 %	21.8 %	16.5 %	22.5 %	22.1 %	32.0 %	23.9 %	23.6 %	26.6 %	21.6 %
		7.8	−1.6	−3.57	−0.47	−4.49	**10.7**	0.32	0.1	0.95	−3.24
	No Revisiting	58.5 %	78.2 %	83.5 %	77.5 %	77.9 %	68.0 %	76.1 %	76.4 %	73.4 %	78.4
		−7.8	1.6	**3.57**	0.47	**4.49**	−10.7	−0.32	−0.1	−0.95	**3.24**
		$\chi^2(9{,}22717) = 197.7$, $p < .0001$									

For both courses, the videos with *JumpSkip(JS)*, *LongPause(LP)* and *SpeedDown(SD)* do not show significance in revisiting behaviors. Interestingly, we find that the videos viewed with *Replay (RP)* and *FrequentPause (FP)* are significantly more likely to be revisited, while less revisiting probabilities are found with the *SpeedUp (SU)* and *Passive(PS)*. In Sect. 5, *RP*, *FP* and *SU* are shown to reflect respectively the highest and lowest subjective difficulties. Therefore, we infer that the video difficulty may confound between the interaction patterns and the probability of video revisiting. However, as other patterns are weaker indicators of the perceived difficulty, the revisiting behaviors may in this case be confounded largely by other factors such as the course intrinsics. For example, the videos with *Inactive (IA)* pattern are significantly more likely to be revisited in the RP and less in the DSP. The potential reasons are hard to identify in this case. In this section we highlight the more general findings that videos with *RP* and *FP* patterns are more likely to be revisited, and more follow-ups of this finding will be discussed in Sect. 8.

7 Student Performance

Students in MOOCs often have diverse background and learning abilities. Depending on their levels, MOOC students may watch video lectures in different ways. For example, we may hypothesize that strong students selectively watch

MOOC videos while weak students spend more time with the learning materials. Our research question in this section is ***How do Strong and Weak students differ in lecture video viewing behaviors?*** The video interaction patterns provide us with a handy tool for diagnosing the students' video behaviors, so our analysis will be based on comparing the strategy of employing the patterns. The analysis would deepen our understanding of how students of different performances use MOOCs, thus proving insights for the instructors to design solutions for helping the weak students.

7.1 Method

The foremost challenge for the analysis in this section is to define *Strong* and *Weak* students. Considering MOOC is an open platform, students have different motives. A great proportion of the students drop out in the early or middle of the courses for various reasons. Even those who watch all the videos do not necessarily aim at obtaining a certificate or completing all the learning activities. This means the students who obtain 0 point in the final score are not necessarily weak in their learning abilities. As mentioned in Sect. 3, the two courses in our datasets do not have exams, and weekly quizzes are the only mean for assessing students. The quizzes for the RP course can be submitted unlimited times, and we have seen many students submitted more than 10 times for a quiz. The consequence is that 82 % of the passed students got certificates of distinction, which is a quite inflated percentage. In order to compare the students who are strong and weak in learning abilities, we take a subset of the data which includes only the students who completed all the assignments. Thus we can maximally believe the remaining students have a similar learning goal, which is to complete the courses. As shown in Table 1, in the DSP course, only 263 (less than 3 %) of the total students finished, and only 23 students got distinction results, whilst the RP course has a 23 % completion rate. In the analysis hereafter, we only analyse the RP students who submitted all of the 6 assignments. To simplify the analysis, the students who obtained 80 % of the total points in their *FIRST* submissions are defined as *Strong*. Otherwise, they are labeled as *Weak*. The subset contains 4555 (86.3 %) of the passed students, of which 35.3 % are *Strong* students. In addition, only the first-time watching patterns are analyzed.

7.2 Result

A video session has an expected probability of 37.6 % to come from *Strong* students under the null hypothesis that the employment of video interaction patterns is independent of students' performance (Table 6). Chi-square test shows that the adoptions of video interaction patterns are significantly different between strong and weak students. Post-hoc residual analysis reveals that strong students tend to interact less with the videos, so the frequency of *HighSpeed(HS)*, *SpeedUp(SU)*, *Passive(PS)* and *Inactive(IA)* sessions are significantly higher. On the other hand, weak students interact more with videos, they use significantly more *SkimSkip(SS)*, *JumpSkip (JS)*, *FrequentPause(FP)* and *LongPause(LP)*.

Table 6. Proportion of video interaction patterns based on students performance

	RP	HS	SU	SS	IA	FP	JS	LP	SD	PS
Strong	38.7 %	46.1 %	39.7 %	31.6 %	35.9 %	32.4 %	33.8 %	35.5 %	38.6 %	39.0 %
	−1.2	**17.3**	**2.1**	−4	−9	−6.6	−3.6	−3.2	0.7	**3.8**
Weak	61.3 %	53.9 %	60.3 %	68.4 %	64.1 %	67.6 %	66.2 %	64.5 %	61.4 %	61.0 %
	1.2	−17.3	−2.1	**4**	**9**	**6.6**	**3.6**	**3.2**	−0.7	−3.8
	$\chi^2(9{,}76094) = 406.3$, $p < .0001$									

Recall that both *FrequentPause(FP)* and *Replay(RP)* reflect the highest video difficulty. Interestingly, the usage of *Replay* pattern is not significantly different between the two student groups, suggesting that replaying behaviors may not discriminate students' performance.

We are more interested in deeply understanding two pausing patterns, since on one hand we observe significantly more video sessions of these patterns, on the other hand, the pauses provide MOOC designers with opportunities to augment the video with supporting information. We have discussed in Sect. 5 that pauses may occur when the presented information is overloaded so that the students require buffer time or external material to understand the content. Weak students are found to use both *FrequentPause(FP)* and *LongPause(LP)* significantly more often than strong students. In order to understand how the weak students adopt the two pausing patterns, we randomly selected 50 video sessions with *FrequentyPause(FP)* pattern and another 50 with *LongPause(LP)* pattern from the weak students' interaction logs and manually examine the situations under which the associated 698 pauses happened. We categorize the pauses by the occasions when the professor was explaining *example codes* (46.7 %), *programing grammar* (12.7 %), *demos* (4.8 %), *theories* (33.8 %) and *others* (2 %)(e.g. in summary). Fisher's exact test shows no significant differences in the categories of pauses in sessions between *FrequentPause(FP)* and *LongPause(LP)* patterns ($p = 0.47$). Nearly half of the pauses occur when example code snippets are shown in the video frame, and more than half of the pauses are related to the presentation of code (code, grammar and demo). This result indicates that the weak students may have significant problems in understanding the code compared to strong student. We need solutions to support the weak students in this regard.

8 Design Insights for MOOC Practitioners

Our analyses in this paper show that MOOC students follow different video interaction patterns while watching lecture videos. The strategy of adopting the patterns may vary for different videos, depending on the students' perceived video difficulty, their capability and whether or not a video is watched for the first time. The analyses of the video interaction patterns presented so far provide the following insights for improving the MOOC learning experiences.

1. Detect the Change of Video Interaction Patterns [Algorithm]. The MOOC platform should detect the changes of video interaction patterns for the students, because such changes may indicate variations in perceived video difficulty. If we detect the students are experiencing difficulty, then proper interventions, such as external materials can be suggested to provide a more personalized learning experience.

2. Provide Quick Access for Revisiting Videos [User Interface]. Since a video session with in-video dropout may have 60 % – 70 % chances to be revisited, it is advisable to provide the students with short-cut access to those videos. In addition, we find out that video sessions with interaction patterns that reflect higher difficulty, such as the *Replay(RP)* and *FrequentPause(FP)* are more likely to be revisited. A possible solution is to design a side bar which lists the videos that will be potentially revisited, so that the students can be aware of them and make easy re-access.

3. Make Use of the Pauses [Information Interaction]. Weak students tend to use more *FrequentPause(FP)* or *LongPause(LP)* to buffer the professors' explanation in the video into their mind. It is advisable to reduce the information overload in the lecture slides. However, this might be less feasible for programming courses, as often code blocks have to be presented. A possible solution is to make use of the time when the pauses occur. For example, when a student pauses at a particular video frame with code blocks, auxiliary information that helps the students to understand the code can be displayed as overlay.

9 Conclusions and Future Work

This paper shed light on the relationships between video interaction patterns and some important aspects in MOOC learning such as the perceived video difficulty, video revisiting behaviors and students' performance. Based on the findings, our research proposed design suggestions to improve MOOC learning experiences. Concrete implementations of these general suggestions as well as validations of their effectiveness are however left as future work. In this work we pursue generalization through statistical inferences rather than specificity, so the patterns were categorized based on in-video interaction features without considering the longitudinal nature of the course or the differences in video content. In addition, MOOC is a multi-faceted learning platform, lecture videos play an important role but do not explain the complete picture. Activities in the forum, quiz, students' motivations all may relate to the aspects we described in this paper. Future work may also incorporate these factors to gain a more comprehensive understanding about how students learn in MOOCs.

References

1. Branch, P., Egan, G., Tonkin, B.: Modeling interactive behaviour of a video based multimedia system. In: 1999 IEEE International Conference on Communications, ICC 1999, vol. 2, pp. 978–982. IEEE (1999)

2. Breslow, L., Pritchard, D.E., DeBoer, J., Stump, G.S., Ho, A.D., Seaton, D.: Studying learning in the worldwide classroom: research into edxs first mooc. Res. Pract. Assess. **8**, 13–25 (2013)
3. Dey-Sircar, J.K., Salehi, J.D., Kurose, J.F., Towsley, D.: Providing vcr capabilities in large-scale video servers. In: Proceedings of the Second ACM International Conference on Multimedia MULTIMEDIA 1994, pp. 25–32. ACM, New York (1994)
4. Dolnicar, S., Grabler, K., Mazanec, J.A., Woodside, A., Crouch, G., Oppermann, M., Sakai, M. et al.: A tale of three cities: perceptual charting for analysing destination images. CABI Publishing (1999)
5. Guo, P.J., Reinecke, K.: Demographic differences in how students navigate through moocs. In: Proceedings of the First ACM Conference on Learning @ Scale Conference L@S 2014, pp. 21–30. ACM, New York (2014)
6. Halawa, S., Greene, D., Mitchell, J.: Dropout prediction in moocs using learner activity features. In: Proceedings of the Second European MOOC Stakeholder Summit, pp. 58–65 (2014)
7. Kim, J., Guo, P.J., Seaton, D.T., Mitros, P., Gajos, K.Z., Miller, R.C.: Understanding in-video dropouts and interaction peaks inonline lecture videos. In: Proceedings of the First ACM Conference on Learning @ Scale Conference L@S 2014, pp. 31–40. ACM, New York (2014)
8. Kizilcec, R.F., Piech, C., Schneider, E.: Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge LAK 2013, pp. 170–179. ACM, New York (2013)
9. Li, N., Kidzinski, L., Jermann, P., Pierre, D.: How do in-video interactions reflect perceived video difficulty? In: Proceedings of the Third European MOOC Stakeholder Summit (to appear, 2015)
10. Li, V.O.K., Liao, W., Qiu, X., Wong, E.W.M.: Performance model of interactive video-on-demand systems. IEEE J. Sel. Areas Commun. **14**(6), 1099–1109 (1996)
11. Mongy, S., Djeraba, C., Simovici, D.A.: On clustering users' behaviors in video sessions. In: Stahlbock, R., Crone, S.F., Lessmann, S. (eds.) DMIN, pp. 99–103. CSREA Press (2007)
12. Seaton, D.T., Bergner, Y., Chuang, I., Mitros, P., Pritchard, D.E.: Who does what in a massive open online course? Commun. ACM **57**(4), 58–65 (2014)
13. Shenoy, P.J., Vin, H.M.: Efficient support for scan operations in video servers. In: Proceedings of the third ACM international conference on Multimedia, pp. 131–140. ACM (1995)
14. Syeda-Mahmood, T., Ponceleon, D.: Learning video browsing behavior and its application in the generation of video previews. In: Proceedings of the Ninth ACM International Conference on Multimedia MULTIMEDIA 2001, pp. 119–128. ACM, New York (2001)
15. Sinha, T., Jermann, P., Li, N., Dillenbourg, P.: Your click decides your fate: inferring information processing and attrition behavior from mooc video clickstream interactions. In: Proceedings of the 2014 Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses (2013)